Time-Varying Smooth Transition Autoregressive Models

Stefan Lundbergh
Department of Economic Statistics, Stockholm School of Economics, Box 6501, SE-113 83 Stockholm, Sweden (*Stefan.Lundbergh@hhs.se*)

Timo Teräsvirta
Department of Economic Statistics, Stockholm School of Economics, Box 6501, SE-113 83 Stockholm, Sweden (*Timo.Terasvirta@hhs.se*)

Dick van Dijk
Econometric Institute, Erasmus University Rotterdam, P.O. Box 1738, NL-3000 DR Rotterdam, The Netherlands (*djvandijk@few.eur.nl*)

Nonlinear regime-switching behavior and structural change are often perceived as competing alternatives to linearity. In this article we study the so-called time-varying smooth transition autoregressive (TV-STAR) model, which can be used both for describing simultaneous nonlinearity and structural change and for distinguishing between these features. Two modeling strategies for empirical specification of TV-STAR models are developed. Monte Carlo simulations show that neither of the two strategies dominates the other. A specific-to-general-to-specific procedure is best suited for obtaining a first impression of the importance of nonlinearity and/or structural change for a particular time series. A specific-to-general procedure is most useful in careful specification of a model with nonlinear and/or time-varying properties. An empirical application to a large dataset of U.S. macroeconomic time series illustrates the relative merits of both modeling strategies.

KEY WORDS: Nonlinearity; Structural change; Time series model specification.

1. INTRODUCTION

Over the years, ample empirical evidence has been gathered for both nonlinearity and structural change in the dynamic properties of many observed time series. For example, the dynamic behavior of macroeconomic time series depends nonlinearly on the phase of the business cycle. This regime-switching behavior related to expansion and contraction periods has been the focus of much research, (see, e.g., Teräsvirta and Anderson 1992; Beaudry and Koop 1993; Sichel 1993; Tiao and Tsay 1994; Potter 1995; Pesaran and Potter 1997). On the other hand, Stock and Watson (1996) reported an overwhelming amount of evidence for instability in both univariate and multivariate models for a large number of U.S. postwar macroeconomic time series.

Despite these indications that both nonlinearity and structural change are relevant for many time series, to date these features have mainly been analyzed in isolation. This dichotomy may be due to how time series modeling usually is carried out. Typically, one starts with a linear model. The estimated linear model is routinely subjected to misspecification tests, including tests for nonlinearity and parameter nonconstancy. If the misspecification tests indicate that the linear model is inadequate, then the model is modified accordingly. Modeling is usually terminated by estimating this alternative model. Thus, when nonlinearity is found and modeled, parameter constancy of the estimated nonlinear model is rarely tested (see, however, Eitrheim and Teräsvirta 1996; Teräsvirta 1998 for exceptions). Conversely, when parameter constancy is rejected in a linear model, the estimated time-varying parameter model is rarely tested for nonlinearity.

It should be mentioned, however, that some work has been done to consider nonlinearity and structural change simultaneously. Diebold and Rudebusch (1992), Watson (1994) and Parker and Rothman (1996) applied nonparametric techniques to examine whether certain characteristics of the business cycle, such as the duration and amplitude of recessions and booms, have changed over time, while allowing these properties to be different in different business cycle phases. Cooper (1998) found indications for both regime-switching behavior and structural change in U.S. industrial production in a regression tree analysis. Some attempts to capture both nonlinearity and structural instability with parametric time series models have been made as well. For example, Lütkepohl, Teräsvirta, and Wolters (1998) and Wolters, Teräsvirta, and Lütkepohl (1999) used smooth transition models to examine linearity and stability of German money demand equations. Skalin and Teräsvirta (2002) applied a model closely related to the one discussed in this article to describe business cycle asymmetry and changing seasonal variation in quarterly unemployment rates of a number of developed countries. Finally, Kim and Nelson (1999) and Luginbuhl and De Vos (1999) allowed for structural change in the mean growth rate of U.S. real gross domestic product (GDP) while modeling dynamic behavior in recessions and expansions by means of a Markov switching model.

Journal of Business & Economic Statistics
January 2003, Vol. 21, No. 1
DOI 10.1198/073500102288618810

In this article, we study a time series model, based on the principle of smooth transition, which allows for nonlinear dynamics in conjunction with time-varying parameters. This time-varying smooth transition autoregressive (TV-STAR) model can be used both for describing simultaneous nonlinearity and structural change and for distinguishing between these two features. The model may be regarded as a special case of the multiple-regime STAR (MRSTAR) model of Van Dijk and Franses (1999). The article is organized as follows. In Section 2 we define the TV-STAR model and discuss its main properties. In Section 3 we describe two different strategies for building TV-STAR models. We investigate the performance of these procedures by Monte Carlo simulation in Section 4. In Section 5 we apply our methodology to test for nonlinearity and structural change in a large number of U.S. macroeconomic time series. An example involving the U.S. help-wanted advertising index illustrates the stages of our modeling strategy in more detail. Finally, we provide some concluding remarks in Section 6.

2. THE MODEL

The smooth transition autoregressive (STAR) model of Teräsvirta (1994) serves as the basic building block for the TV-STAR model. The STAR model for a univariate time series y_t, which is observed at $t = 1-p, -p, \dots, -1, 0, 1, \dots, T-1, T$, is given by

$$y_t = \boldsymbol{\varphi}_1'\mathbf{x}_t(1 - G(s_t; \gamma, c)) + \boldsymbol{\varphi}_2'\mathbf{x}_t G(s_t; \gamma, c) + \varepsilon_t, \quad (1)$$

where $\mathbf{x}_t = (1, \tilde{\mathbf{x}}_t')'$ with $\tilde{\mathbf{x}}_t = (y_{t-1}, \dots, y_{t-p})'$, $\boldsymbol{\varphi}_i = (\varphi_{i,0}, \varphi_{i,1}, \dots, \varphi_{i,p})'$, $i = 1, 2$, and ε_t is a white noise process. It is straightforward to extend the model to allow for exogenous variables $z_{1t}, \dots, z_{kt}$ as additional regressors. The resultant smooth transition regression (STR) model was discussed in detail by Teräsvirta (1998).

In general, the so-called transition function $G(s_t; \gamma, c)$ in (1) is a continuous function that is bounded between 0 to 1. To simplify the exposition, here we restrict attention to the logistic function

$$G(s_t; \gamma, c) = [1 + \exp\{-\gamma(s_t - c)\}]^{-1} \qquad \text{with } \gamma > 0. \quad (2)$$

It is straightforward to modify the analysis presented herein to allow for different transition functions. In (2), s_t is the transition variable, and γ and c are slope and location parameters. The parameter restriction $\gamma > 0$ is an identifying restriction. As s_t increases, the logistic function changes monotonically from 0 to 1, with the change being symmetric around c, as $G(c - z; \gamma, c) = 1 - G(c + z; \gamma, c)$ for all z. As $\gamma \to \infty$, the logistic function $G(s_t; \gamma, c)$ approaches the indicator function $\mathrm{I}[s_t > c]$. Finally, for $\gamma = 0$, $G(s_t; \gamma, c) = 1/2$ for all s_t, such that the model reduces to a linear dynamic model.

The STAR model (1) with (2) may be thought of as a regime-switching model that allows for two regimes, associated with the extreme values of the transition function, $G(s_t; \gamma, c) = 0$ and $G(s_t; \gamma, c) = 1$, where the transition from one regime to the other is smooth. The regime that occurs at time t is determined by the observable variable s_t. Teräsvirta (1994) focused on the case where $s_t = y_{t-d}$ is a lagged dependent variable. Lin and Teräsvirta (1994) considered an AR model with smoothly time-varying parameters (TV-AR), which is obtained by setting $s_t = t$. In this article we combine the STAR and TV-AR models, with the idea to make it possible to model continuous structural change and smooth transition-type nonlinearity simultaneously. The resultant time-varying STAR (TV-STAR) model is given by

$$\begin{aligned} y_t = {} & [\boldsymbol{\varphi}_1'\mathbf{x}_t(1 - G(y_{t-d})) + \boldsymbol{\varphi}_2'\mathbf{x}_t G(y_{t-d})][1 - G(t)] \\ & + [\boldsymbol{\varphi}_3'\mathbf{x}_t(1 - G(y_{t-d})) + \boldsymbol{\varphi}_4'\mathbf{x}_t G(y_{t-d})]G(t) + \varepsilon_t, \quad (3) \end{aligned}$$

where $G(y_{t-d}) = G(y_{t-d}; \gamma_1, c_1)$ and $G(t) = G(t; \gamma_2, c_2)$ are logistic functions defined as in (2). The TV-STAR model can be interpreted as describing y_t by a STAR model at all times, with a smooth change in the autoregressive parameters, from $\boldsymbol{\varphi}_1$ and $\boldsymbol{\varphi}_2$ to $\boldsymbol{\varphi}_3$ and $\boldsymbol{\varphi}_4$ in the regimes corresponding with $G(y_{t-d}) = 0$ and $G(y_{t-d}) = 1$. This can be seen from the alternative representation of (3),

$$y_t = \boldsymbol{\varphi}_1(t)'\mathbf{x}_t(1 - G(y_{t-d})) + \boldsymbol{\varphi}_2(t)'\mathbf{x}_t G(y_{t-d}) + \varepsilon_t, \quad (4)$$

where

$$\boldsymbol{\varphi}_1(t) = \boldsymbol{\varphi}_1[1 - G(t)] + \boldsymbol{\varphi}_3 G(t) \quad (5)$$

and

$$\boldsymbol{\varphi}_2(t) = \boldsymbol{\varphi}_2[1 - G(t)] + \boldsymbol{\varphi}_4 G(t). \quad (6)$$

Equations (4), (5), and (6) form the alternative hypothesis in tests of parameter constancy of a STAR model (see Eitrheim and Teräsvirta 1996; Teräsvirta 1998). Model (3) may also be viewed as a special case of the multiple-regime STAR model of Van Dijk and Franses (1999).

Submodels of the general TV-STAR model in (3) can be obtained by imposing appropriate restrictions on the autoregressive parameters in the different regimes. In particular, if $\boldsymbol{\varphi}_1 \neq \boldsymbol{\varphi}_2$, $\boldsymbol{\varphi}_1 = \boldsymbol{\varphi}_3$, and $\boldsymbol{\varphi}_2 = \boldsymbol{\varphi}_4$ in (3), then the resultant model is a two-regime STAR model. Similarly, if $\boldsymbol{\varphi}_1 = \boldsymbol{\varphi}_2$, $\boldsymbol{\varphi}_3 = \boldsymbol{\varphi}_4$, and $\boldsymbol{\varphi}_1 \neq \boldsymbol{\varphi}_3$, then the model reduces to a TV-AR model. Models that are nested within the general TV-STAR model are considered in more detail in Section 4.

3. SPECIFICATION OF TV-STAR MODELS

Applying the TV-STAR model to data requires a coherent modeling strategy. It must be decided whether or not the full TV-STAR model as given in (3) is required, or whether the time series under consideration may be adequately characterized by a submodel, such as a STAR or TV-AR model. Either way, there are other data-based choices to be made, such as selecting the delay parameter d and the lag length p. Doing all this in a systematic fashion is highly desirable.

Granger (1993) strongly recommended a "specific-to-general" strategy for building nonlinear time series models. This implies starting with a simple or restricted model and proceeding to more complicated models only if diagnostic tests indicate that the maintained model is inadequate. In the present situation, an additional (statistical) motivation for adopting such an approach is that the identification problems discussed herein prevent us from starting with a full TV-STAR model and reducing its size by, say, a sequence of likelihood ratio tests.

Choosing the specific-to-general approach still leaves two possibilities. First, the procedure proposed by Teräsvirta (1994) for two-regime STAR models can be extended to obtain a "specific-to-general" modeling cycle for TV-STAR models. This approach consists of several specification, estimation, and evaluation stages, starting with a linear AR model and proceeding toward the full TV-STAR model via a STAR or TV-AR model. Second, given the particular choice of transition variables in the TV-STAR model, a more direct "specific-to-general-to-specific" approach also seems feasible. This procedure begins with testing linearity directly against the TV-STAR model (specific-to-general). If linearity is rejected, then subhypotheses are tested to determine whether a STAR model or a TV-AR model provides an adequate characterization of the time series at hand (general-to-specific).

3.1 Lagrange Multiplier Tests

Both modeling procedures described here make heavy use of Lagrange multiplier (LM) tests. For the sake of completeness, we illustrate the form of these test statistics by means of an example.

Consider the problem of testing the null hypothesis of linearity against the alternative of a TV-STAR model as given in (3). To obtain an appropriate test statistic, it is useful to rewrite the model as

$$y_t = \boldsymbol{\alpha}'\mathbf{x}_t + \boldsymbol{\beta}'\mathbf{x}_t G^*(y_{t-d}) + \boldsymbol{\pi}'\mathbf{x}_t G^*(t) + \boldsymbol{\theta}'\mathbf{x}_t G^*(y_{t-d})G^*(t) + \varepsilon_t, \quad (7)$$

where $\boldsymbol{\alpha}$, $\boldsymbol{\beta}$, $\boldsymbol{\pi}$, and $\boldsymbol{\theta}$ are linear combinations of $\boldsymbol{\varphi}_1, \dots, \boldsymbol{\varphi}_4$, and $G^*(s_t) = G(s_t) - 1/2$. This reparameterization allows us to specify the null hypothesis of linearity as $\mathsf{H}_0 : \gamma_1 = \gamma_2 = 0$. We make the following assumptions about the model in (7):

(A.1) The errors $\varepsilon_t \sim \text{NID}(0, \sigma^2)$, $t = 1, \dots, T$. (This assumption is required only for the definition of the likelihood function and can be relaxed.)

(A.2) Under H_0, $1 - \sum_{j=1}^{p} \alpha_j L^j \neq 0$ for all $|L| \leq 1$, and furthermore, $\mathsf{E}[y_t^4] < \infty$.

As indicated earlier, model (7) is identified only under the alternative hypothesis. In particular, under the null hypothesis, $\boldsymbol{\beta}$, $\boldsymbol{\pi}$, and $\boldsymbol{\theta}$ are unidentified nuisance parameters, which renders standard asymptotic inference invalid (see Davies 1977, 1987; Hansen 1996). We circumvent the identification problem of Luukkonen, Saikkonen, and Teräsvirta (1988) by approximating the two transition functions in (7) by a first-order Taylor expansion around the null hypothesis. The resulting auxiliary regression is

$$y_t = \boldsymbol{\alpha}^{*\prime}\mathbf{x}_t + \boldsymbol{\beta}^{*\prime}\tilde{\mathbf{x}}_t y_{t-d} + \boldsymbol{\pi}^{*\prime}\mathbf{x}_t t + \boldsymbol{\theta}^{*\prime}\tilde{\mathbf{x}}_t y_{t-d} t + R(\gamma_1, \gamma_2) + \varepsilon_t. \quad (8)$$

The term $R(\gamma_1, \gamma_2)$ represents the remainder from the two Taylor expansions. Under H_0, we have $R(\gamma_1, \gamma_2) = 0$, so this remainder does not affect our asymptotic distribution theory. In going from (7) to (8), we have implicitly assumed that $d \leq p$; that is, y_{t-d} is an element of $\tilde{\mathbf{x}}_t$. If this is not the case, then $\beta_0^* y_{t-d}$ must be added as auxiliary regressor to (8).

In (8), $\boldsymbol{\alpha}^* = \boldsymbol{\alpha}$ and $\boldsymbol{\beta}^* = \boldsymbol{\pi}^* = \boldsymbol{\theta}^* = \mathbf{0}$ if and only if $\gamma_1 = \gamma_2 = 0$. Hence the null hypothesis of linearity in the auxiliary regression (8) is given by $\mathsf{H}_0 : \boldsymbol{\beta}^* = \boldsymbol{\pi}^* = \boldsymbol{\theta}^* = \mathbf{0}$, which can be tested by means of an LM test in a straightforward manner. Under assumptions (A.1) and (A.2), the test statistic, denoted as $LM_{\text{TV-STAR}}$ in the following, has an asymptotic chi-squared distribution with $3p+1$ degrees of freedom under H_0. The test can be computed as follows:

1. Regress y_t on $\mathbf{x}_t$ and compute the sum of squared residuals, SSR_0.
2. Regress y_t on $\mathbf{x}_t$, $\tilde{\mathbf{x}}_t y_{t-d}$, $\mathbf{x}_t t$, and $\tilde{\mathbf{x}}_t y_{t-d} t$ and compute the sum of squared residuals, SSR_1.
3. Compute the chi-squared version of the test statistic as $LM_{\text{TV-STAR}} = \frac{T(\text{SSR}_0 - \text{SSR}_1)}{\text{SSR}_0}$, or the F version as $LM_{\text{TV-STAR}} = \frac{(\text{SSR}_0 - \text{SSR}_1)/(3p+1)}{\text{SSR}_1/(T-4p-2)}$, where T denotes the sample size.

The F version of the test is recommended in small and moderate samples. This is because the asymptotically correct chi-squared test may be severely oversized in small samples if the dimension of the null hypothesis is large, which may often be the case here. On the other hand, the size of the F test is close to the nominal size even in small samples and, at the same time, the test has reasonable power. Simulations in the STAR case (Teräsvirta and Anderson 1992) support this conclusion.

Special cases of the auxiliary regression (8) yield other linearity tests. It follows from the expressions of the parameters $\boldsymbol{\beta}^*$, $\boldsymbol{\pi}^*$, and $\boldsymbol{\theta}^*$ in terms of the parameters in the original TV-STAR model that (a) $\boldsymbol{\beta}^* = \mathbf{0}$ if and only if $\gamma_1 = 0$, (b) $\boldsymbol{\pi}^* = \mathbf{0}$ if and only if $\gamma_2 = 0$, and (c) $\boldsymbol{\theta}^* = \mathbf{0}$ if $\gamma_1 = 0$ or $\gamma_2 = 0$, or both. Hence, assuming $\boldsymbol{\pi}^* = \boldsymbol{\theta}^* = 0$ in (8), linearity can be tested against STAR by testing $\mathsf{H}_0 : \boldsymbol{\beta}^* = \mathbf{0}$. Similarly, linearity can be tested against the alternative of smoothly time-varying parameters by assuming $\boldsymbol{\beta}^* = \boldsymbol{\theta}^* = \mathbf{0}$ and testing $\mathsf{H}_0 : \boldsymbol{\pi}^* = \mathbf{0}$. The linearity tests against STAR and TV-AR of Teräsvirta (1994) and Lin and Teräsvirta (1994) are thus special cases of the test presented here.

Finally, the LM-type test derived earlier is based on linearization of the TV-STAR model. Hence some information about the form of the nonlinear structure under the alternative is lost, which may adversely affect the power of the tests. As an alternative, one might consider the tests of Andrews (1993) and Andrews and Ploberger (1994), using the simulation-based technique developed by Hansen (1996) to obtain critical values. However, as pointed out by Skalin (1998), Hansen's approach would be computationally very demanding for tests against STAR (or TV-AR or TV-STAR), because the model under the alternative hypothesis contains at least two unidentified nuisance parameters. Furthermore, the small-sample results of Hansen (1996) indicate that the LM-type tests can compare quite favorably with the optimal tests of Andrews (1993) and Andrews and Ploberger (1994) in terms of size and power.

3.2 Specific-to-General Approach

The specific-to-general modeling cycle for TV-STAR models can be summarized as follows (see Van Dijk and Franses 1999 for a description of a similar procedure for specification of multiple-regime STAR models):

1. Specify an AR model for the time series of interest y_t.
2. Test the null hypothesis of linearity separately against smoothly changing parameters [$\mathsf{H}_0 : \boldsymbol{\pi}^* = \mathbf{0}$, assuming $\boldsymbol{\beta}^* = \boldsymbol{\theta}^* = \mathbf{0}$ in (8)] and against STAR ($\mathsf{H}_0 : \boldsymbol{\beta}^* = \mathbf{0}$, assuming $\boldsymbol{\pi}^* = \boldsymbol{\theta}^* = \mathbf{0}$). The appropriate value of the delay parameter d in the alternative STAR model can be determined by carrying out the latter test for y_{t-d} with $d = 1, \ldots, d_{\max}$ as a transition variable, and selecting the value of d for which the p value of the statistic is smallest (cf. Teräsvirta 1994). If the null is not rejected against either alternative, then retain the AR model.
3. Estimate the model under the alternative for which the null hypothesis is rejected most strongly (measured in terms of p value) and compute LM-type tests against additional nonlinear or time-varying structure. For example, if linearity is rejected most convincingly against STAR nonlinearity, then estimate a two-regime STAR model and test its parameter constancy against the alternative of smoothly changing parameters. Such misspecification tests were developed by Eitrheim and Teräsvirta (1996) to evaluate estimated STAR models. These test statistics can be easily modified to diagnose a model with smoothly changing parameters.
4. Estimate a TV-STAR model if the null of no remaining nonlinearity or parameter constancy is rejected in step 3. Evaluate the model by computing misspecification tests of no remaining nonlinearity and parameter constancy, using generalizations of the tests of Eitrheim and Teräsvirta (1996) for evaluating STAR models. Otherwise, tentatively accept the null model.

3.3 Specific-to-General-to-Specific Approach

Although the specific-to-general procedure makes intuitive sense, it has some drawbacks. First, it may involve the estimation of several nonlinear models. Another, more important concern is that the form of the final model is conditional on the path selected. We may reach a TV-STAR model both through a STAR and through a TV-AR model. It is likely that these two paths lead to different final models. The first path emphasizes nonlinearity in the series, and the model is completed with nonconstant parameters. The second path stresses parameter nonconstancy, and the variation not explained by this assumption is assigned to nonlinearity. If the aim is to test economic theory, then this ambivalence may appear worrisome. If the main aim is forecasting, then the model builder may want to retain both models and, for example, combine forecasts from them into a final forecast.

Another complication that may arise in the specific-to-general approach is caused by the fact that the tests of linearity against STAR and TV-AR used in step 2 are not robust to structural change and to nonlinearity. Hence these tests cannot discriminate perfectly between these competing alternatives. In fact, it may happen quite often that linearity is rejected against both STAR and TV-AR, leaving the researcher with the decision in which direction to proceed. In the Monte Carlo experiments in Section 4, we investigate whether the suggested rule to proceed in the direction for which the null hypothesis is rejected most convincingly (step 3) is helpful in making this decision.

An alternative modeling approach is to adopt a specific-to-general-to-specific procedure, which starts with testing linearity directly against the TV-STAR model. If the null hypothesis is rejected, then one checks whether a STAR or a TV-AR model captures the essential features of the time series under investigation. The resulting modeling cycle consists of the following steps:

1. Specify an AR model for the time series of interest y_t.
2. Use the $LM_{\text{TV-STAR}}$ statistic described in Section 3.1 to test linearity directly against the TV-STAR alternative [$\mathsf{H}_0^{\text{TV-STAR}} : \boldsymbol{\beta}^* = \boldsymbol{\pi}^* = \boldsymbol{\theta}^* = \mathbf{0}$ in (8)]. The appropriate value of the delay parameter d can be determined by computing the $LM_{\text{TV-STAR}}$ statistic for y_{t-d} with $d = 1, \ldots, d_{\max}$ as a transition variable, and selecting the value of d for which the p value of the statistic is smallest.
3. If the null hypothesis of linearity is rejected, then test subhypotheses that are nested in $\mathsf{H}_0^{\text{TV-STAR}}$ to assess whether a TV-STAR model is really necessary to characterize the time series y_t or whether either a STAR model or a TV-AR model is sufficient. In particular, test

$$\mathsf{H}_0^{\text{STAR}}: \boldsymbol{\beta}^* = \boldsymbol{\theta}^* = \mathbf{0}$$

and

$$\mathsf{H}_0^{\text{TV-AR}}: \boldsymbol{\pi}^* = \boldsymbol{\theta}^* = \mathbf{0}$$

in the auxiliary regression (8), for the value of d selected in step 2. The corresponding LM statistics, which are denoted as LM_{STAR} and $LM_{\text{TV-AR}}$, have asymptotic chi-squared distributions with $2p$ and $2p+1$ degrees of freedom. From the relationships between $\boldsymbol{\beta}^*$, $\boldsymbol{\pi}^*$, and $\boldsymbol{\theta}^*$ and the parameters γ_1 and γ_2, it follows that under $\mathsf{H}_0^{\text{STAR}}$, the model reduces to a TV-AR model, whereas under $\mathsf{H}_0^{\text{TV-AR}}$, a STAR model results. These considerations lead to the following decision rule:

- If both $\mathsf{H}_0^{\text{STAR}}$ and $\mathsf{H}_0^{\text{TV-AR}}$ are rejected, then retain the TV-STAR model.
- If $\mathsf{H}_0^{\text{STAR}}$ is rejected but $\mathsf{H}_0^{\text{TV-AR}}$ is not, then select a STAR model.
- If $\mathsf{H}_0^{\text{TV-AR}}$ is rejected but $\mathsf{H}_0^{\text{STAR}}$ is not, then select a TV-AR model.

The one combination of outcomes that does not imply a clear-cut choice is when neither $\mathsf{H}_0^{\text{STAR}}$ nor $\mathsf{H}_0^{\text{TV-AR}}$ is rejected but the general null hypothesis $\mathsf{H}_0^{\text{TV-STAR}}$ is rejected. In this case, one may resort to the LM-type tests, which test linearity against STAR and against TV-AR separately as in step 2 of the specific-to-general approach. In this way one can find out which (if any) of the submodels of the TV-STAR model is most suitable for the time series at hand.

4. Estimate the selected model and evaluate it by misspecification tests.

This strategy has two drawbacks. First, testing directly against the full TV-STAR model may imply weak power, because the dimension of the null hypothesis may be large in many cases. Second, as a referee pointed out, testing subhypotheses within the auxiliary regression (8) may not be appropriate, because the original null hypothesis $H_0^{\text{TV-STAR}}$ has already been rejected. In other words, the remainder in (8) may not equal 0 under the null hypotheses H_0^{STAR} and $H_0^{\text{TV-AR}}$, which may affect the size of the corresponding tests. Thus the additional tests may be seen just as additional model selection devices that are likely to be helpful if the difference in p values of the two tests is reasonably large. Note, however, that a similar strategy works well in choosing between logistic and exponential STAR models (see Teräsvirta 1994 for a discussion).

3.4 Estimation and Evaluation of TV-STAR Models

The parameters of a TV-STAR model may be estimated by nonlinear least squares (NLS). When assumption (A.1) is valid, NLS is equivalent to maximum likelihood (based on a Gaussian likelihood function). If assumption (A.1) does not hold, then the NLS estimates can be interpreted as quasi-maximum likelihood estimates. Under suitable regularity conditions, (e.g., conditions M1–M7 of Wooldridge 1994, pp. 2653–2655), the NLS estimates are consistent and asymptotically normal. Issues that deserve particular attention, (and that we discuss next) include concentrating the sum of squares function, choosing starting values, and estimating the smoothness parameters γ_1 and γ_2.

When the parameters in the transition functions $G(y_{t-d}; \gamma_1, c_1)$ and $G(t; \gamma_2, c_2)$ are known, the TV-STAR model is linear in the remaining AR parameters. Thus, conditional on $\boldsymbol{\gamma} = (\gamma_1, \gamma_2)'$ and $\mathbf{c} = (c_1, c_2)'$, estimates of $\boldsymbol{\varphi} = (\boldsymbol{\varphi}_1', \boldsymbol{\varphi}_2', \boldsymbol{\varphi}_3', \boldsymbol{\varphi}_4')'$ can be obtained by ordinary least squares (OLS) as

$$\hat{\boldsymbol{\varphi}}(\boldsymbol{\gamma}, \mathbf{c}) = \left(\sum_{t=1}^{T} \mathbf{x}_t(\boldsymbol{\gamma}, \mathbf{c})\mathbf{x}_t(\boldsymbol{\gamma}, \mathbf{c})'\right)^{-1} \left(\sum_{t=1}^{T} \mathbf{x}_t(\boldsymbol{\gamma}, \mathbf{c}) y_t\right), \quad (9)$$

where $\mathbf{x}_t(\boldsymbol{\gamma}, \mathbf{c}) = (\mathbf{x}_t'(1 - G(y_{t-d}))(1 - G(t)), \mathbf{x}_t' G(y_{t-d})(1 - G(t)), \mathbf{x}_t'(1 - G(y_{t-d}))G(t), \mathbf{x}_t' G(y_{t-d})G(t))'$, and the notation $\hat{\boldsymbol{\varphi}}(\boldsymbol{\gamma}, \mathbf{c})$ is used to indicate that the estimate of $\boldsymbol{\varphi}$ is conditional on $\boldsymbol{\gamma}$ and $\mathbf{c}$. As suggested by Leybourne, Newbold, and Vougas (1998), this implies that the sum of squares function can be concentrated with respect to $\boldsymbol{\varphi}_1$, $\boldsymbol{\varphi}_2$, $\boldsymbol{\varphi}_3$, and $\boldsymbol{\varphi}_4$ as

$$Q_T(\boldsymbol{\gamma}, \mathbf{c}) = \sum_{t=1}^{T} (y_t - \hat{\boldsymbol{\varphi}}(\boldsymbol{\gamma}, \mathbf{c})' \mathbf{x}_t(\boldsymbol{\gamma}, \mathbf{c}))^2.$$

This reduces the computational burden considerably, because $Q_T(\boldsymbol{\gamma}, \mathbf{c})$ need only be minimized with respect to the four parameters γ_1, c_1, γ_2, and c_2. The corresponding estimates of $\boldsymbol{\varphi}$ follow from (9) at each iteration of the nonlinear optimization procedure.

The foregoing also suggests that sensible starting values for the nonlinear optimization can be obtained by performing a grid search over γ_1, c_1, γ_2, and c_2 and selecting the parameter values that minimize the sum of squared residuals $Q_T(\boldsymbol{\gamma}, \mathbf{c})$. Making γ_1 and γ_2 scale-free, as discussed by Teräsvirta (1994), helps in determining a useful set of grid values for these parameters.

The remarks of Teräsvirta (1994, 1998) concerning the apparent inaccuracy of the estimate of the smoothness parameter γ in ST(A)R models when γ is large apply to the TV-STAR model as well. We therefore just repeat the fact that in such a situation, low "t values" for the estimates of γ_1 and γ_2 cannot be interpreted as evidence against the TV-STAR model (for details, see Teräsvirta 1994, 1998).

4. PERFORMANCE OF MODELING STRATEGIES

In this section we evaluate the small-sample performance of the two modeling strategies for TV-STAR models by Monte Carlo simulation. First, we assess the size and power properties of the LM-type tests used in the specific-to-general-to-specific approach in Section 4.2. The size and power properties of the LM-type tests used in the specific-to-general procedure have been investigated elsewhere (see Teräsvirta 1994; Lin and Teräsvirta 1994; Eitrheim and Teräsvirta 1996). Second, in Section 4.3 we examine the "success rate" of the two procedures, that is, the frequency of selecting the correct model when it is included in the set of alternatives.

4.1 Monte Carlo Design

We examine the properties of the modeling strategies for seven different data-generating processes (DGPs), all of which are nested in the TV-STAR model (3) with $p = d = 1$,

$$\begin{aligned} y_t = {} & [\varphi_1 y_{t-1}(1 - G(y_{t-1})) + \varphi_2 y_{t-1} G(y_{t-1})][1 - G(t)] \\ & + [\varphi_3 y_{t-1}(1 - G(y_{t-1})) + \varphi_4 y_{t-1} G(y_{t-1})]G(t) + \varepsilon_t, \end{aligned} \quad (10)$$

and $\varepsilon_t \sim \text{NID}(0, \sigma^2)$, where we set $\sigma^2 = 1$. In all of the experiments described herein we use 10,000 replications and an effective sample size of 250 observations. Necessary starting values for the artificial time series are set equal to 0, whereafter the first 100 observations are discarded to eliminate any potential influence of this choice. Finally, the autoregressive order p and the delay parameter d are assumed known throughout.

The DGPs can be conveniently characterized by the restrictions that they impose on the autoregressive parameters in (10) as follows:

(a) $\varphi_1 = \varphi_2 = \varphi_3 = \varphi_4$. In this case the TV-STAR model (10) reduces to a linear AR(1) model. This DGP is used only to evaluate the size properties of the LM-type tests in the specific-to-general-to-specific approach. The AR parameter is varied among $\varphi_1 = \{0, .1, .3, \ldots, .9\}$.

(b) $\varphi_1 = \varphi_3$ and $\varphi_2 = \varphi_4$. Under these restrictions, the DGP is a STAR model with AR parameters φ_1 and φ_2 in the regimes corresponding with $G(y_{t-1}) = 0$ and $G(y_{t-1}) = 1$. The parameterizations for this DGP are taken from Luukkonen et al. (1988). We set φ_1 equal to $-.5$ or $.5$, whereas φ_2 is varied among $\varphi_2 \in \{-.9, -.7, -.5, -.3, 0, .3, .5, .7, .9\}$. For the logistic function $G(y_{t-1})$, we set $\gamma_1 = 5$ and $c_1 = 0$.

(c) $\varphi_1 = \varphi_2$ and $\varphi_3 = \varphi_4$. In this case the DGP is a TV-AR model. Again, φ_1 is fixed at $-.5$ or $.5$, whereas φ_3 is varied as φ_2 in DGP (b). To examine the impact of the timing of the parameter change, the location parameter c_2 in $G(t)$ is varied among $c_2 = \{.25, .50, .75\}$. Furthermore, we set $\gamma_2 = 25$, which in this case implies that the parameter change takes about half of the sample to be completed.

(d) $\varphi_1 = \varphi_2$. The DGP resulting from this restriction is a linear AR model that evolves into a STAR model as $G(t)$ increases from 0 to 1.

(e) $\varphi_1 = \varphi_3$. This DGP is a STAR model with a constant AR parameter in the regime $G(y_{t-1}) = 0$, whereas the AR parameter in the regime $G(y_{t-1}) = 1$ changes smoothly from φ_2 to φ_4.

(f) $\varphi_4 - \varphi_3 = \varphi_2 - \varphi_1$. This DGP is a TV-STAR model where the (absolute) difference between the AR parameters in the two regimes of the effective STAR model (which in a sense can be interpreted as the degree of nonlinearity of the model) remains constant over time. That is, the difference between $\boldsymbol{\varphi}_1(t)$ and $\boldsymbol{\varphi}_2(t)$ as given in (5) and (6) is the same for all $0 \le t \le 1$.

(g) $\varphi_1 = \varphi_4$. This restriction renders a TV-STAR model in which the dynamic behavior in the regime $G(y_{t-1}) = 0$ before the structural change is the same as the dynamic behavior in the regime $G(y_{t-1}) = 1$ after the change.

DGPs (d), (e) and (g) restrict two of the four AR parameters $\varphi_j, j = 1, \ldots, 4$. In all cases, these restricted parameters are set equal to .5, whereas the two unrestricted parameters are varied independently among $\{-.9, -.7, -.5, -.3, 0, .3, .5, .7, .9\}$. In DGP (f), we fix $\varphi_1 = .5$ and vary φ_2 and φ_4 independently as in the other DGPs. For each combination of these unrestricted parameters, the value of φ_3 is obtained as $\varphi_3 = \varphi_1 - \varphi_2 + \varphi_4$. We investigate only those parameter combinations that satisfy sufficient conditions for weak stationary. Configurations with $|\varphi_3| \ge 1$ are thus not considered. Finally, in DGPs (d)–(g), the values of the slope and location parameters in the transition functions $G(y_{t-1})$ and $G(t)$ are defined as in DGPs (b) and (c).

4.2 Size and Power of LM-Type Statistics

In this section we consider the small-sample properties of the three LM-type test statistics involved in the specific-to-general-to-specific procedure, as discussed in Section 3.3. Rejection frequencies at nominal significance levels $\alpha = .01$, .05, and .10 when the DGP is an AR(1) model are given in Table 1. It is seen that the empirical size of all three tests is reasonably close to the selected nominal significance levels, although the LM_{STAR} statistic becomes quite conservative as φ_1 approaches 1.

Detailed results for the power experiments involving a STAR model [DGP (b)] and a TV-AR model [DGP (c)] are not given here, but they are available from the authors on request. The rejection frequencies for the tests under these DGPs behave as expected. For example, in case of DGP (b), the rejection frequencies of the $LM_{\text{TV-STAR}}$ and LM_{STAR} statistics are reasonably high and increase monotonically as the degree of nonlinearity (as measured by the absolute difference between the AR parameters in the two regimes $|\varphi_2 - \varphi_1|$) increases. The rejection frequencies of the $LM_{\text{TV-AR}}$ statistic are close to the nominal significance level for all values of φ_2, as expected. Under DGP (c), the rejection frequencies of the LM_{STAR} statistic are close to the nominal significance level for all values of φ_3, whereas the power of the $LM_{\text{TV-STAR}}$ and $LM_{\text{TV-AR}}$ statistics increases monotonically as the difference $|\varphi_1 - \varphi_3|$ increases. The latter two tests are most powerful if the change in parameters is centered around the middle of the sample.

For DGPs (d)–(g), which involve both nonlinearity and structural change, the behavior of the three statistics also corresponds to expectations. We illustrate this by presenting some results for DGPs (d) and (e). Detailed results for DGPs (f) and (g) are available from the authors on request. The graphs in the left column of Figure 1 show the rejection frequencies of the three statistics at the 5% nominal significance level in the case where the DGP is a linear model that changes into a STAR model [DGP (d)], with the change centered at the middle of the sample ($c_2 = .50$).

The power of the $LM_{\text{TV-STAR}}$ and $LM_{\text{TV-AR}}$ tests is seen to increase as the difference between φ_3 and φ_4 and the restricted parameters φ_1 and φ_2 (which are fixed at .5) becomes larger. In contrast, the power of the LM_{STAR} statistic increases when the difference between φ_3 and φ_4 (the AR parameters in the STAR model after the change) becomes larger. Comparing these results with the findings for the same DGP but with $c_2 = .25$ and $c_2 = .75$ shows that the power of the $LM_{\text{TV-STAR}}$ and $LM_{\text{TV-AR}}$ statistics is highest when the change from the AR to the STAR model is centered at the middle of the sample, whereas the power of the LM_{STAR} statistic is highest when this change occurs earlier in the sample.

The rejection frequencies for the various tests in the case where the DGP is a STAR model with a smoothly changing AR parameter in the regime corresponding to $G(y_{t-1}) = 1$

Table 1. Empirical Size of LM-Type Tests in the Specific-to-General-to-Specific Procedure

		$LM_{TV\text{-}STAR}$			LM_{STAR}			$LM_{TV\text{-}AR}$		
φ_1	α	.010	.050	.100	.010	.050	.100	.010	.050	.100
0		.009	.042	.089	.009	.046	.095	.009	.043	.088
.1		.008	.041	.087	.008	.044	.090	.010	.040	.085
.3		.008	.039	.085	.007	.042	.086	.009	.040	.085
.5		.007	.039	.086	.006	.040	.079	.009	.041	.082
.7		.006	.035	.078	.004	.031	.067	.007	.040	.084
.9		.010	.041	.084	.003	.018	.045	.013	.055	.100

NOTE: Empirical size of the LM-type test statistics that are involved in the specific-to-general-to-specific procedure for specification of TV-STAR models, as described in Section 3.3. Series are generated according to an AR(1) model with autoregressive parameter φ_1. The table is based on 10,000 replications.

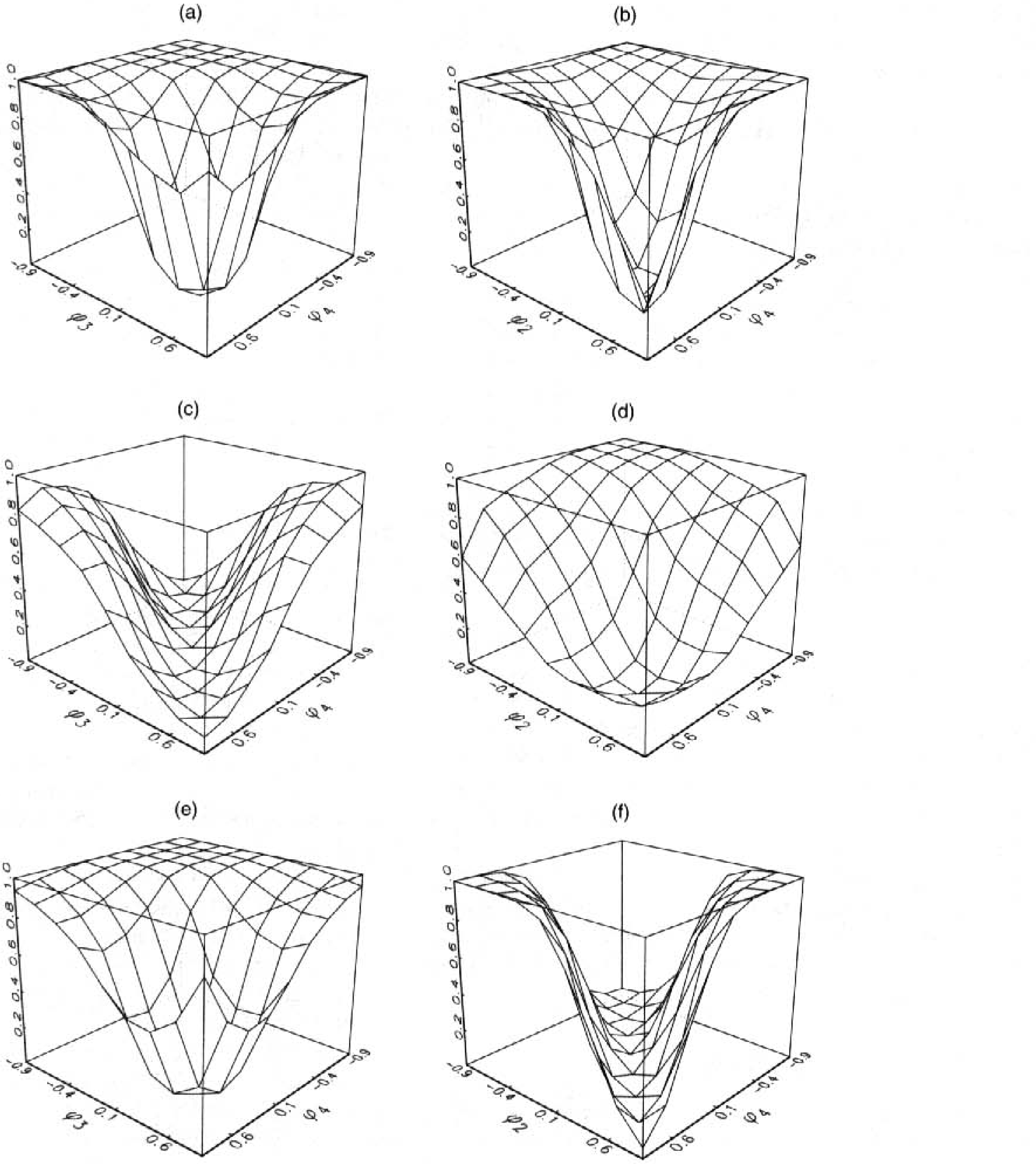

Figure 1. Rejection Frequencies of the Appropriate Null Hypotheses by the Various LM-Type Tests that Are Part of the Specific-to-General-to-Specific Procedure Outlined in Section 3.3. Results are shown for F variants of the tests at the 5% nominal significance level. Artificial series are generated according to DGP (d) (left column) or DGP (e) (right column) with $c_2 = .50$, as described in Section 4.1. The graphs are based on 10,000 replications. (a) DGP (d), $LM_{\text{TV-STAR}}$; (b) DGP (e), $LM_{\text{TV-STAR}}$; (c) DGP (d), LM_{STAR}; (d) DGP (e), LM_{STAR}; (e) DGP (d), $LM_{\text{TV-AR}}$; (f) DGP (e), $LM_{\text{TV-AR}}$.

only [DGP (e)] are shown in the right column of Figure 1. Again, we present results only for $c_2 = .50$. The power of the $LM_{\text{TV-STAR}}$ statistic is seen to be an increasing function of the absolute differences between the AR parameters in the two STAR regimes before and after the change, that is, $|\varphi_2 - \varphi_1|$ and $|\varphi_4 - \varphi_3|$. Changing the parameter c_2 to .25 or .75 affects the power of this statistic in obvious ways. For example, if the change occurs early in the sample, then the difference $|\varphi_4 - \varphi_3|$ is much more important than the difference $|\varphi_2 - \varphi_1|$. This holds even more so for the power of the LM_{STAR} statistic. In the case where $c_2 = .25$, the power of this test is determined almost entirely by the value of φ_4 (relative to φ_3), whereas the value of φ_2 hardly seems to matter. The power of the $LM_{\text{TV-AR}}$ statistic increases as the absolute difference between φ_4 and φ_2 increases, which is also as expected.

4.3 Simulating Model Selection Strategies

Next we examine the selection frequencies of the various models when using the model specification strategies outlined in Sections 3.2 and 3.3. Throughout, we use a nominal significance level of 5% to determine the significance of the (diagnostic) test statistics at the various stages of the specific-to-general procedure and the $LM_{\text{TV-STAR}}$, LM_{STAR}, and

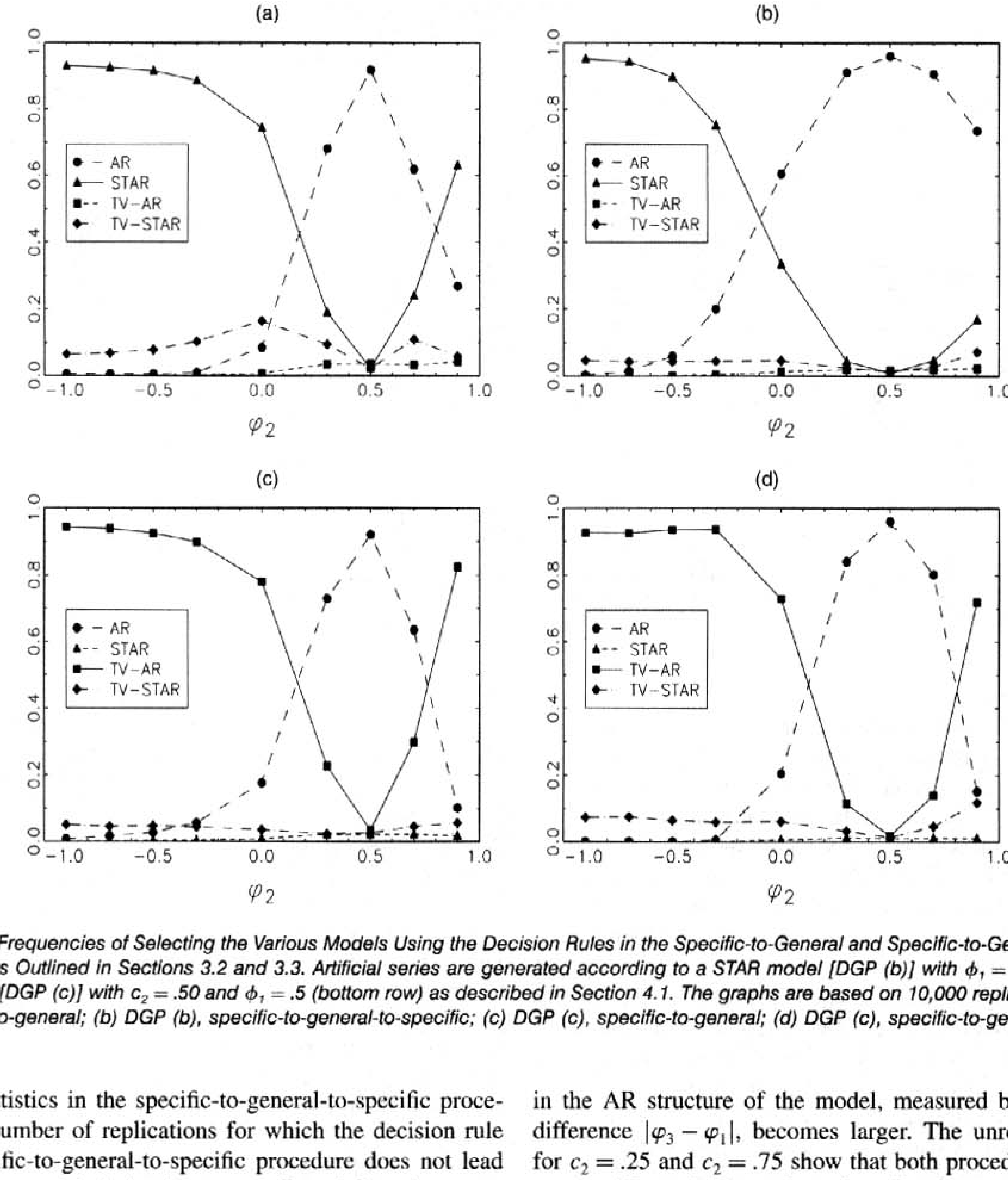

Figure 2. Frequencies of Selecting the Various Models Using the Decision Rules in the Specific-to-General and Specific-to-General-to-Specific Procedures as Outlined in Sections 3.2 and 3.3. Artificial series are generated according to a STAR model [DGP (b)] with $\phi_1 = .5$ (top row) or a TV-AR model [DGP (c)] with $c_2 = .50$ and $\phi_1 = .5$ (bottom row) as described in Section 4.1. The graphs are based on 10,000 replications. (a) DGP (b), specific-to-general; (b) DGP (b), specific-to-general-to-specific; (c) DGP (c), specific-to-general; (d) DGP (c), specific-to-general-to-specific.

$LM_{\text{TV-AR}}$ statistics in the specific-to-general-to-specific procedure. The number of replications for which the decision rule in the specific-to-general-to-specific procedure does not lead to a clear-cut model choice is very small, and thus those outcomes are not reported separately. To save space, in the following we discuss only selected (but representative) results in detail; a complete set of simulation results is available from the authors on request.

The frequencies of selecting a given model by the specific-to-general and the specific-to-general-to-specific procedures in the case where the DGP is a STAR model [DGP (b)] with $\phi_1 = .5$ are displayed in the top row of Figure 2. Clearly, the correct model (STAR) is selected more frequently as the absolute difference between the AR parameters in the two regimes, $|\varphi_2 - \varphi_1|$, increases. The specific-to-general procedure performs slightly better than the specific-to-general-to-specific procedure, especially for small and moderate values of the difference $|\varphi_2 - \varphi_1|$.

The frequencies of selecting the different models for series generated according to a TV-AR model [DGP (c)] with $c_2 = .50$ and $\phi_1 = .5$ are given in the bottom row of Figure 2. The correct model (TV-AR) is chosen more often as the change in the AR structure of the model, measured by the absolute difference $|\varphi_3 - \varphi_1|$, becomes larger. The unreported results for $c_2 = .25$ and $c_2 = .75$ show that both procedures select the correct AR model less often when the change occurs early or late in the sample. Again, the true model is selected slightly more frequently by the specific-to-general procedure than by the specific-to-general-to-specific procedure.

Finally, Table 2 contains model selection frequencies for DGP (g). Both procedures tend to select the TV-STAR model as the differences between the AR parameters in the two regimes of the STAR models before and after the parameter change are sufficiently large. The specific-to-general-to-specific procedure selects the true model more frequently than the specific-to-general procedure if $|\varphi_2 - \varphi_1|$ and $|\varphi_4 - \varphi_3|$ are large, whereas the reverse holds for more subtle nonlinearities and structural changes.

The results from simulating the remaining DGPs (available on request) can be roughly summarized as follows. When the DGP is a TV-STAR model, the specific-to-general-to-specific strategy yields the best results. When the DGP is a submodel, either STAR or TV-AR, the specific-to-general strategy yields the correct model more frequently.

Table 2. Model Selection Frequencies, DGP (g)

		Specific-to-general				Specific-to-general-to-specific			
ϕ_2	ϕ_3	AR	STAR	TV-AR	TV-STAR	AR	STAR	TV-AR	TV-STAR
−.9	−.9	.290	.001	.016	.693	0	0	.017	.983
−.9	−.5	.456	.003	.031	.509	0	0	.048	.952
−.9	0	.530	.019	.053	.399	.001	0	.113	.886
−.9	.5	.189	.118	.062	.631	.010	.007	.119	.863
−.9	.9	.005	.093	.091	.810	0	.003	.041	.956
−.5	−.5	.398	.002	.032	.569	0	0	.135	.865
−.5	0	.636	.018	.062	.284	.023	.003	.284	.688
−.5	.5	.311	.178	.115	.397	.099	.037	.268	.590
−.5	.9	.008	.099	.155	.739	.003	.019	.141	.833
0	0	.679	.007	.052	.261	.296	.015	.368	.307
0	.3	.793	.049	.078	.079	.568	.028	.232	.156
0	.5	.610	.161	.140	.088	.613	.052	.193	.121
0	.7	.230	.239	.217	.315	.391	.104	.279	.167
0	.9	.017	.118	.308	.557	.043	.105	.378	.426
.5	.5	.919	.022	.033	.026	.961	.009	.017	.008
.5	.7	.690	.057	.117	.136	.858	.028	.067	.017
.5	.9	.105	.132	.386	.377	.244	.116	.386	.077
.9	.9	.727	.025	.112	.135	.398	.019	.396	.144

NOTE: Frequencies of selecting the various models using the decision rules in the specific-to-general and specific-to-general-to-specific procedures as outlined in Sections 3.2 and 3.3. Series are generated according to DGP (g) with $c_2 = .50$, as described in Section 4.1. The table is based on 10,000 replications.

The general conclusion is, not unexpectedly, that neither of the two strategies dominates the other. The specific-to-general-to-specific procedure has the advantage in that no nonlinear estimation is required at the specification stage. This may also be regarded as a weakness, because at first sight it could appear natural to estimate a TV-STAR model if linearity is rejected against it, and thereafter try to reduce it to either submodel by specification tests. This alternative is not directly available, however, due to the aforementioned identification problems. It would only be possible to estimate either a STAR or TV-AR model and test its adequacy against TV-STAR. That step is in fact a part of the specific-to-general strategy. In practice, the investigator may well want to test linearity against STAR, TV-AR, *and* TV-STAR and use all of the evidence from these tests in model selection. The relative merits of the two strategies are discussed further in the next section.

5. NONLINEARITY AND STRUCTURAL INSTABILITY IN U.S. MACROECONOMIC TIME SERIES

5.1 Testing for Nonlinearity and Structural Instability

As discussed in Section 1, both (business cycle) nonlinearity and structural instability have been found to be important features of macroeconomic time series. However, most of the evidence for nonlinearity has been obtained under the assumption of parameter constancy, whereas structural change has been detected using mainly linear models. In this section we use the TV-STAR framework to consider the joint presence of nonlinearity and instability in macroeconomic time series.

We examine the dataset compiled by Stock and Watson (1999), consisting of 214 monthly U.S. macroeconomic time series. The series are grouped in the categories that appear in Tables 3 and 4, with the number of series in each category in parentheses. The sample period starts in January 1959 (although some series are not available from the beginning) and ends in December 1996, a total of 456 observations. The time series are transformed as was done by Stock and Watson (1999); that is, series in dollars, real quantities, and price deflators are transformed to logarithms. Many of the time series are seasonally adjusted. A detailed description of the dataset appears in the appendix of Stock and Watson (1999).

Stock and Watson (1999) assessed the importance of nonlinearity for these time series by investigating whether STAR models and artificial neural networks produce forecasts that improve on those from linear models. Stock and Watson (1996) investigated the presence of structural change in a comparable (but smaller) dataset using various tests for parameter constancy. We examine nonlinearity and structural change simultaneously using the TV-STAR framework. Specifically, we apply the LM-type tests and associated model selection rule from the specific-to-general-to-specific procedure.

An issue of ongoing debate about macroeconomic time series variables is whether they are best characterized as trend stationary or as difference stationary. We apply the procedure under both assumptions in models for levels and first differences of the series. This also allows us to examine the sensitivity of the results regarding the presence of nonlinearity and structural change in the time series to the assumptions made about their long-run properties. Under the assumption of trend stationarity, we first extract a linear deterministic trend from the data by estimating the regression

$$y_t = \alpha + \beta t + u_t,$$

and subsequently examine the residuals $\hat{u}_t$ for nonlinearity and structural change.

The order of the AR model under the null hypothesis is determined by the Akaike information criterion (AIC), with the maximum order set equal to $p_{\max} = 18$. Because remaining residual autocorrelation may be mistaken for nonlinearity, (see Teräsvirta 1994), we apply the Ljung–Box test to examine the significance of the first 12 residual autocorrelations in

Table 3. Model Selection for U.S. Macroeconomic Time Series Using Specific-to-General-to-Specific Procedure—Trend Stationarity

Group	AR	STAR	TV-AR	TV-STAR	NLAR
Production (24)	14(19)	3(2)	3(2)	4(1)	0(0)
(Un)Employment (29)	18(22)	5(2)	1(2)	4(1)	1(2)
Wages and salaries (7)	6(6)	0(0)	1(1)	0(0)	0(0)
Construction (21)	17(19)	3(1)	0(0)	1(1)	0(0)
Trade (10)	8(7)	0(2)	1(1)	1(0)	0(0)
Inventories (10)	5(8)	1(0)	4(2)	0(0)	0(0)
Orders (14)	12(14)	2(0)	0(0)	0(0)	0(0)
Money and credit (20)	10(14)	4(1)	1(3)	3(1)	2(1)
Stock returns (11)	10(11)	1(0)	0(0)	0(0)	0(0)
Dividends and volume (3)	1(3)	0(0)	0(0)	2(0)	0(0)
Interest rates (11)	2(7)	7(3)	0(1)	2(0)	0(0)
Exchange rates (6)	5(5)	0(0)	0(0)	1(1)	0(0)
Producer price inflation (16)	11(15)	0(0)	1(0)	3(0)	1(1)
Consumer price inflation (16)	6(15)	8(0)	1(1)	0(0)	1(0)
Consumption (5)	1(2)	0(0)	2(2)	2(1)	0(0)
Miscellaneous (11)	3(7)	4(1)	2(3)	2(0)	0(0)
Total (214)	129(174)	38(12)	17(18)	25(6)	5(4)

NOTE: Number of time series for which the different models are selected using the decision rule in the specific-to-general-to-specific procedure outlined in Section 3.3, under the assumption of trend stationarity. The numbers in parentheses relate to the procedure in case heteroscedasticity robust versions of the LM-type tests are used. The "NLAR" column contains the number of series for which the outcomes of the LM tests in the specific-to-general-to-specific procedure are in conflict and cannot be used to select an appropriate model.

the AR(p) model selected by the AIC. If necessary, the lag length p is increased until the null hypothesis of no residual autocorrelation can no longer be rejected at the 5% significance level. The final AR order differs from the order selected by AIC for 70 series under trend stationarity and 37 series under difference stationarity. Besides residual autocorrelation, neglected heteroscedasticity may also lead to spurious rejection of the null hypothesis. We therefore also report results based on robustified tests as developed by Wooldridge (1991).

As we consider macroeconomic time series, we are interested mainly in nonlinearity related to the business cycle. The transition variable in the (TV-)STAR model should therefore reflect the property that expansion and contraction regimes are sustained periods of growth and decline. This excludes monthly changes, because they are too noisy to be reliable indicators of the business cycle regime (see Birchenhall, Jessen, Osborn and Simpson 1999; Skalin and Teräsvirta, 2002). For that reason, under the assumption of difference stationarity, we use 12-month differences as a transition variable, that is, $s_t = \Delta_{12} y_{t-d} \equiv y_{t-d} - y_{t-d-12}$, $d = 1, \ldots, d_{\max}$, with the maximum value of the delay parameter $d_{\max} = 6$. Under the assumption of trend stationarity, we consider lagged deviations from the deterministic trend, $\hat{u}_{t-d}$, as a transition variable, again with $d_{\max} = 6$. In both cases, the value of d for which the p value of the $LM_{\text{TV-STAR}}$ statistic is smallest is selected as appropriate. The LM_{STAR} and $LM_{\text{TV-AR}}$ statistics

Table 4. Model Selection for U.S. Macroeconomic Time Series Using Specific-to-General-to-Specific Procedure—Difference Stationarity

Group	AR	STAR	TV-AR	TV-STAR	NLAR
Production (24)	5(17)	12(2)	2(1)	4(0)	1(4)
(Un)Employment (29)	14(24)	9(1)	2(2)	4(1)	0(1)
Wages and salaries (7)	6(7)	0(0)	0(0)	0(0)	1(0)
Construction (21)	12(14)	8(4)	0(0)	1(1)	0(2)
Trade (10)	6(8)	3(1)	0(1)	1(0)	0(0)
Inventories (10)	6(9)	1(0)	0(0)	1(0)	2(1)
Orders (14)	8(8)	5(4)	0(0)	1(0)	0(2)
Money and credit (20)	12(15)	2(1)	1(0)	4(4)	1(0)
Stock returns (11)	11(10)	0(1)	0(0)	0(0)	0(0)
Dividends and volume (3)	2(3)	0(0)	0(0)	1(0)	0(0)
Interest rates (11)	0(10)	8(0)	0(0)	3(1)	0(0)
Exchange rates (6)	6(6)	0(0)	0(0)	0(0)	0(0)
Producer price inflation (16)	7(15)	3(0)	0(0)	6(0)	0(1)
Consumer price inflation (16)	2(10)	3(3)	2(0)	4(0)	5(3)
Consumption (5)	0(2)	0(0)	0(2)	4(0)	1(1)
Miscellaneous (11)	3(9)	3(0)	0(1)	3(1)	2(0)
Total (214)	100(167)	57(17)	7(7)	37(8)	13(15)

NOTE: Number of U.S. macroeconomic time series for which the different models are selected using the decision rule in the specific-to-general-to-specific procedure outlined in Section 3.3, under the assumption of difference stationarity. The numbers in parentheses relate to the procedure in case heteroscedasticity robust versions of the LM-type tests are used. The "NLAR" column contains the number of series for which the outcomes of the LM tests in the specific-to-general-to-specific procedure are in conflict and cannot be used to select an appropriate model.

are computed only for this choice of transition variable and are used to decide on the appropriate model using the model selection rule, where a nominal 5% significance level is used throughout. As mentioned in Section 3.3, there is a combination of test results for which the specific-to-general-to-specific procedure does not lead to a clear-cut choice of the appropriate model. The series for which this indeterminacy occurs are recorded separately in Tables 3 and 4 under the heading "NLAR" (nonlinear AR).

When using standard versions of the LM-type tests, a nonlinear model is selected for 40% of the series under trend stationarity and 53% of the series under difference stationarity. For 12% and 17% of the series a TV-STAR model is the preferred choice. This suggests that both nonlinearity and structural change are present in only a small number of these macroeconomic time series. Most evidence for nonlinearity is found for series in the production, (un)employment, interest rates, and producer and consumer price inflation groups. Nonlinear models appear to be least promising for stock returns and exchange rates.

Note that a TV-AR model is selected for only very few series, especially under the difference stationarity assumption. This finding is in apparent disagreement with the results of Stock and Watson (1996, 1999) indicating that structural instability is far more important than nonlinearity for macroeconomic time series. Nevertheless, the argument in these articles is based on tests for structural change only (which convincingly reject parameter constancy in linear models) and point forecasts from STAR models (which do not improve on point forecasts from linear models). Thus it seems that the issue is not settled: different methods yield different results.

The evidence for nonlinearity and structural change is reduced considerably when robustified versions of the LM tests are used. A nonlinear model is selected only in 19% of the series under trend stationarity and 22% of the series under difference stationarity. The change in test outcomes is especially apparent for groups such as production, (un)employment, interest rates, and inflation. This may suggest that heteroscedasticity is an important feature of many time series in the dataset and should be taken into account. However, there is also the possibility that robustifying the linearity tests against heteroscedasticity reduces their power. We leave this question for further research.

5.2 An Illustrated Example: Help-Wanted Advertising Index

In this section we apply the specific-to-general procedure to one of the time series in the dataset, the help-wanted advertising index (LHELX). This index is based on counts of the number of help-wanted advertisements published in the classified section of newspapers in 51 major American cities (see Abraham 1987; and Zagorsky 1998). It is constructed as the ratio of the number of help-wanted ads to the number of unemployed people in the civilian labor force. The help-wanted advertising index is often used as a proxy for the job-vacancy rate, an important indicator of labor demand and the business cycle. This interpretation of the help-wanted index must be used with caution, however. Abraham (1987) pointed out that the index has been drifting upward relative to the underlying job vacancy variable. This drift may be due to the shift in the composition of vacancies (away from blue-collar jobs toward white-collar jobs, which in general are more heavily advertised), changes in employer advertising practices (particularly due to increased equal employment opportunity and affirmative action pressures), and the decline in the number of competing newspapers in metropolitan areas. Because of these changes, one cannot exclude the possibility of parameter nonconstancy in a time series model for the help-wanted index.

The help-wanted index series is shown in Figure 3. It appears that the series displays asymmetric cyclical behavior, characterized by steep declines during business cycle recessions followed by slow(er) increases during expansions.

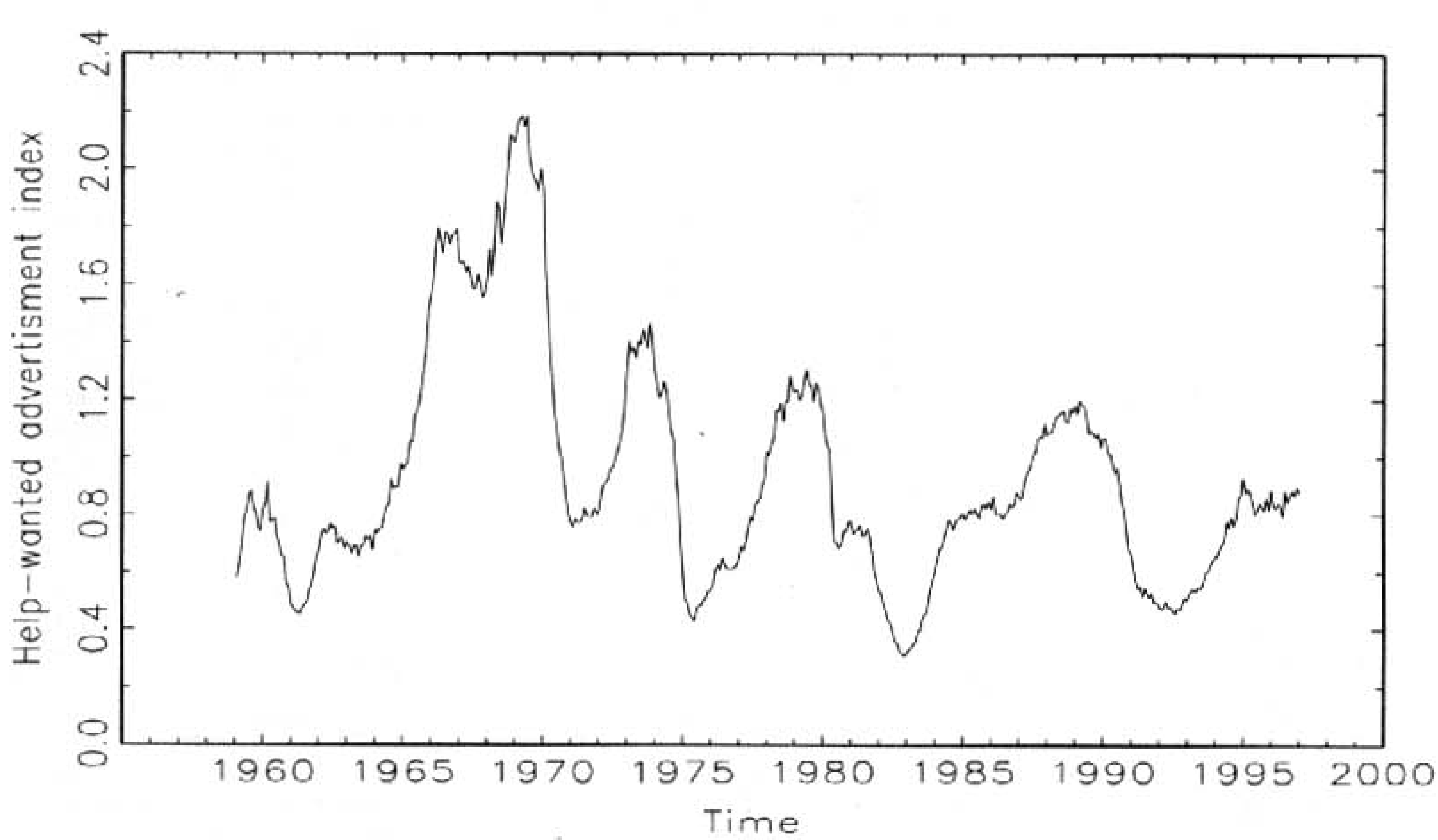

Figure 3. Monthly Help-Wanted Advertisement Index, January 1959–December 1996.

Table 5. Diagnostic Tests for Estimated Models for the Help-Wanted Index

		AR	TV-AR	TV-STAR
$T\hat{\sigma}^2_\varepsilon$		.775	.744	.691
AIC		−6.307	−6.316	−6.377
SIC		−6.251	−6.195	−6.227
SK		$-.52(4.9\times10^{-6})$	$-.54(2.3\times10^{-6})$	$-.43(1.4\times10^{-4})$
EK		$4.00(1.1\times10^{-65})$	$3.96(3.0\times10^{-64})$	$4.02(2.2\times10^{-72})$
LJB		$311(2.6\times10^{-68})$	$306(3.5\times10^{-67})$	$336(1.4\times10^{-73})$
ARCH(1)		$8.94(2.7\times10^{-3})$	$10.00(1.6\times10^{-3})$	$11.22(8.1\times10^{-4})$
ARCH(4)		$39.20(6.3\times10^{-8})$	$43.81(7.0\times10^{-9})$	$24.23(7.2\times10^{-5})$
$LM_{SC}(4)$	S	1.27(.28)	1.56(.18)	1.58(.18)
	R	1.06(.38)	1.01(.40)	1.50(.20)
$LM_{SC}(8)$	S	1.45(.18)	1.00(.43)	1.04(.40)
	R	.93(.49)	.96(.47)	.99(.44)

NOTE: The table presents diagnostic tests for the estimated AR, TV-AR, and TV-STAR models for the help-wanted index over the sample period July 1960–December 1996. T denotes the effective sample size ($T = 437$), $\hat{\sigma}^2_\varepsilon$ is the residual variance, SK is skewness, EK is excess kurtosis, LJB is the Lomnicki–Jarque–Bera test of normality of the residuals, ARCH is the LM test of no autoregressive conditional heteroscedasticity (ARCH), and $LM_{SC}(q)$ denotes (the F variant of) standard (S) and heteroscedasticity-robust (R) versions of the LM test of no serial correlation in the residuals up to and including order q. The numbers in parentheses after the values of the test statistics are p values.

We assume that the help-wanted index series is globally stationary but possibly nonlinear and locally nonstationary and consider models for the level of the series with 12-month differences as a transition variable. Skalin and Teräsvirta (2002) showed, in the context of unemployment rates, how this type of model can generate stationary time series with asymmetric behavior, but such that the series show strong persistence of shocks and appear nonstationary in the light of standard unit root tests. To emphasize this feature, we parameterize the models in terms of first differences, with a lagged level term included as additional regressor.

We begin modeling the LHELX series by specifying a linear AR model. The AIC indicates that a model with four lagged first differences is appropriate, and the Ljung–Box test of no residual autocorrelation up to lag 12 does not reject the null hypothesis at the 5% significance level. To save space, the estimated model is not presented here, but some diagnostic tests are shown in the second column of Table 5.

The linear model appears to have several shortcomings, because the residuals suffer from skewness, excess kurtosis, and heteroscedasticity. Closer inspection of the residuals shows that the apparent nonnormality is due almost entirely to large negative residuals in January 1970 and April 1980 and a large positive residual in April 1968. We apply both modeling strategies for TV-STAR models and begin with the appropriate LM-type tests. Table 6 presents results for standard and heteroscedasticity-robust versions of the test statistics in the specific-to-general-to-specific procedure. It is seen that the p values of the $LM_{\text{TV-STAR}}$ statistic are quite small for all choices of the delay parameter considered here, with the minimum attained for $d = 1$. For $d = 1$ and 2, both null hypotheses associated with the LM_{STAR} and $LM_{\text{TV-AR}}$ statistics can be rejected at conventional significance levels, so that a TV-STAR model would be selected for these values of the delay parameter. Note that the p values of the robustified test statistics are actually smaller than the p values of the standard tests here.

The second and third columns of Table 7, labeled "AR," contain results for the standard and robustified LM-type tests in step 2 of the specific-to-general procedure. Based on the standard tests, linearity can be rejected at the 5% significance level against STAR with $s_t = \Delta_{12}y_{t-d}$ for $d = 1, 2$ as well as against TV-AR ($s_t = t$). The minimum p value is attained against the latter alternative, whereas the robustified tests only allow linearity to be rejected against TV-AR.

Based on the combined test results in the two modeling strategies, we proceed by estimating a TV-AR model. Again, to save space, we do not present parameter estimates for this model here, but the third column of Table 5 contains some summary statistics and results of misspecification tests.

Columns 4 and 5 in Table 7 contain p values for the LM-type tests against additional nonlinear structure. The TV-AR model is tested against the TV-STAR alternative (3) with

Table 6. LM Tests in Specific-to-General-to-Specific Procedure for the Help-Wanted Index

Transition variable	Standard			Robustified		
	$LM_{\text{TV-STAR}}$	LM_{STAR}	$LM_{\text{TV-AR}}$	$LM_{\text{TV-STAR}}$	LM_{STAR}	$LM_{\text{TV-AR}}$
$\Delta_{12}y_{t-1}$	.002	.013	.006	.003	.014	.001
$\Delta_{12}y_{t-2}$	.006	.047	.025	.009	.041	.004
$\Delta_{12}y_{t-3}$	.023	.165	.044	.021	.066	.007
$\Delta_{12}y_{t-4}$	.050	.328	.040	.037	.118	.013
$\Delta_{12}y_{t-5}$	.065	.405	.022	.031	.082	.006
$\Delta_{12}y_{t-6}$	.059	.372	.013	.045	.089	.011

NOTE: p values of LM test statistics in the specific-to-general-to-specific procedure for first differences of the help-wanted index time series, based on an AR(4) model with lagged level term under the null hypothesis.

Table 7. LM (Misspecification) Tests in the Specific-to-General Procedure for the Help-Wanted Index

Transition variable	AR Standard	AR Robustified	TV-AR Standard	TV-AR Robustified	TV-STAR Standard	TV-STAR Robustified
$\Delta_{12}y_{t-1}$	.040	.309	.083	.130	.254	.371
$\Delta_{12}y_{t-2}$	.039	.299	.093	.174	.363	.352
$\Delta_{12}y_{t-3}$	.107	.406	.246	.263	.628	.422
$\Delta_{12}y_{t-4}$	.316	.519	.560	.354	.455	.303
$\Delta_{12}y_{t-5}$	.679	.766	.682	.312	.381	.314
$\Delta_{12}y_{t-6}$	.848	.870	.620	.259	.350	.353
t	.017	.034	.203	.268	.536	.522

NOTE: Columns 2 and 3 contain p values of the LM test statistic in step 2 of the specific-to-general procedure for the help-wanted index time series, based on an AR(4) model for the first differences, which includes a lagged level term. Columns 4 and 5 contain p values of the LM test for remaining nonlinear structure in an estimated TV-AR model. Columns 6 and 7 contain p values of the LM test for remaining nonlinear structure in the estimated TV-STAR model (11)–(13).

transition variable given by $\Delta_{12}y_{t-d}$, $d = 1, \dots, 6$. Some indications for additional structure are found, because the p values for the standard tests with $\Delta_{12}y_{t-d}$ for $d = 1$ and 2 are quite small. This corresponds with the test results in the specific-to-general-to-specific procedure in Table 6, where a TV-STAR model is the preferred choice for these values of d. Using diagnostic tests for the estimated TV-AR model based on higher-order Taylor expansions gives (much) smaller p values, especially when $\Delta_{12}y_{t-1}$ is considered as a transition variable (results are not shown here, but are available on request). Hence we proceed with estimating a TV-STAR model with this choice of transition variable. After sequentially omitting regressors with smallest absolute values of the t statistic until all remaining parameter estimates have absolute t values exceeding 1, we arrive at the model

$$\begin{aligned}
\Delta y_t = {} & [(\underset{\substack{(.120)\\ [.212]}}{.538}\ \Delta y_{t-1} + \underset{\substack{(.115)\\ [.230]}}{.176}\ \Delta y_{t-2}) \times (1 - G(\Delta_{12}y_{t-1})) \\
& + (\underset{\substack{(.011)\\ [.013]}}{.024} - \underset{\substack{(.009)\\ [.013]}}{.017} y_{t-1} + \underset{\substack{(.094)\\ [.147]}}{.135}\ \Delta y_{t-1} + \underset{\substack{(.080)\\ [.119]}}{.148}\ \Delta y_{t-3} \\
& + \underset{\substack{(.089)\\ [.143]}}{.391} \Delta y_{t-4}) \times G(\Delta_{12}y_{t-1})][1 - G(t)] \\
& + \big[(\underset{\substack{(.025)\\ [.020]}}{.037} - \underset{\substack{(.040)\\ [.035]}}{.088} y_{t-1} + \underset{\substack{(.226)\\ [.182]}}{.459}\ \Delta y_{t-3} - \underset{\substack{(.232)\\ [.184]}}{.256}\ \Delta y_{t-4}) \\
& \quad \times (1 - G(\Delta_{12}y_{t-1})) \\
& + (\underset{\substack{(.019)\\ [.013]}}{.037} - \underset{\substack{(.021)\\ [.015]}}{.033} y_{t-1} - \underset{\substack{(.195)\\ [.181]}}{.341}\ \Delta y_{t-1} + \underset{\substack{(.147)\\ [.122]}}{.180}\ \Delta y_{t-2}) \\
& \quad \times G(\Delta_{12}y_{t-1})\big] G(t) + \hat{\varepsilon}_t,
\end{aligned} \tag{11}$$

$$\begin{aligned}
& G(\Delta_{12}y_{t-1}) \\
& = (1 + \exp\{-\underset{\substack{(3.46)\\ [3.10]}}{6.04}(\Delta_{12}y_{t-1} + \underset{\substack{(.041)\\ [.038]}}{.094})/\sigma_{\Delta_{12}y_{t-1}}\})^{-1},
\end{aligned} \tag{12}$$

and

$$G(t) = (1 + \exp\{-\underset{\substack{(3.59)\\ [6.07]}}{7.07}\ (t/T - \underset{\substack{(.097)\\ [.110]}}{.533})\})^{-1}, \tag{13}$$

where ordinary and heteroscedasticity-robust standard errors are given below the parameter estimates in parentheses and brackets.

Results of misspecification tests for the TV-STAR model are given in the rightmost column of Table 5. The two rightmost columns of Table 7 contains p values of LM tests for additional nonlinearity or structural change. The large p values suggest that the foregoing TV-STAR model adequately captures all nonlinearity and instability in the series.

The estimate of the location parameter c_1 is fairly close to 0, indicating that the regimes where $G(\Delta_{12}y_{t-1}) = 0$ and $G(\Delta_{12}y_{t-1}) = 1$ are characterized by positive and negative changes in the series over the past 12 months. Because $G(\Delta_{12}y_{t-1})$ is a monotonic transformation of $\Delta_{12}y_{t-1}$, the periods in which it takes values close to 0 (1) roughly correspond with business cycle recessions (expansions) (see the upper panel of Fig. 4). The transition between these two regimes is smooth, as can be seen in Figure 4(b). The structural change is centered around $t/T = .53$, corresponding to December 1979. The change is very smooth and is not entirely completed by the end of the sample period, as shown in the lower panel of Figure 4(c).

To gain a better understanding of the dynamic properties of the estimated TV-STAR model, it is useful to consider the skeleton of the model, that is, the deterministic part of (11). If the extrapolation of the skeleton is started before the structural change takes place $[G(t) = 0]$ and is carried out without changing the value of $G(t)$, the realizations converge to a unique and stable equilibrium $y_t^* = 1.467$. If the extrapolation is started assuming that $G(t) = 1$, the realizations converge to a limit cycle of 113 months, as shown in Figure 5. The cycle mimics the dynamic properties of the time series during the latter part of the sample period. The range and periodicity of the cycle correspond quite closely with those observed in the empirical time series. It also contains asymmetry, as the parts of the cycle during which the series increases and decreases are somewhat different in length (63 and 50 months).

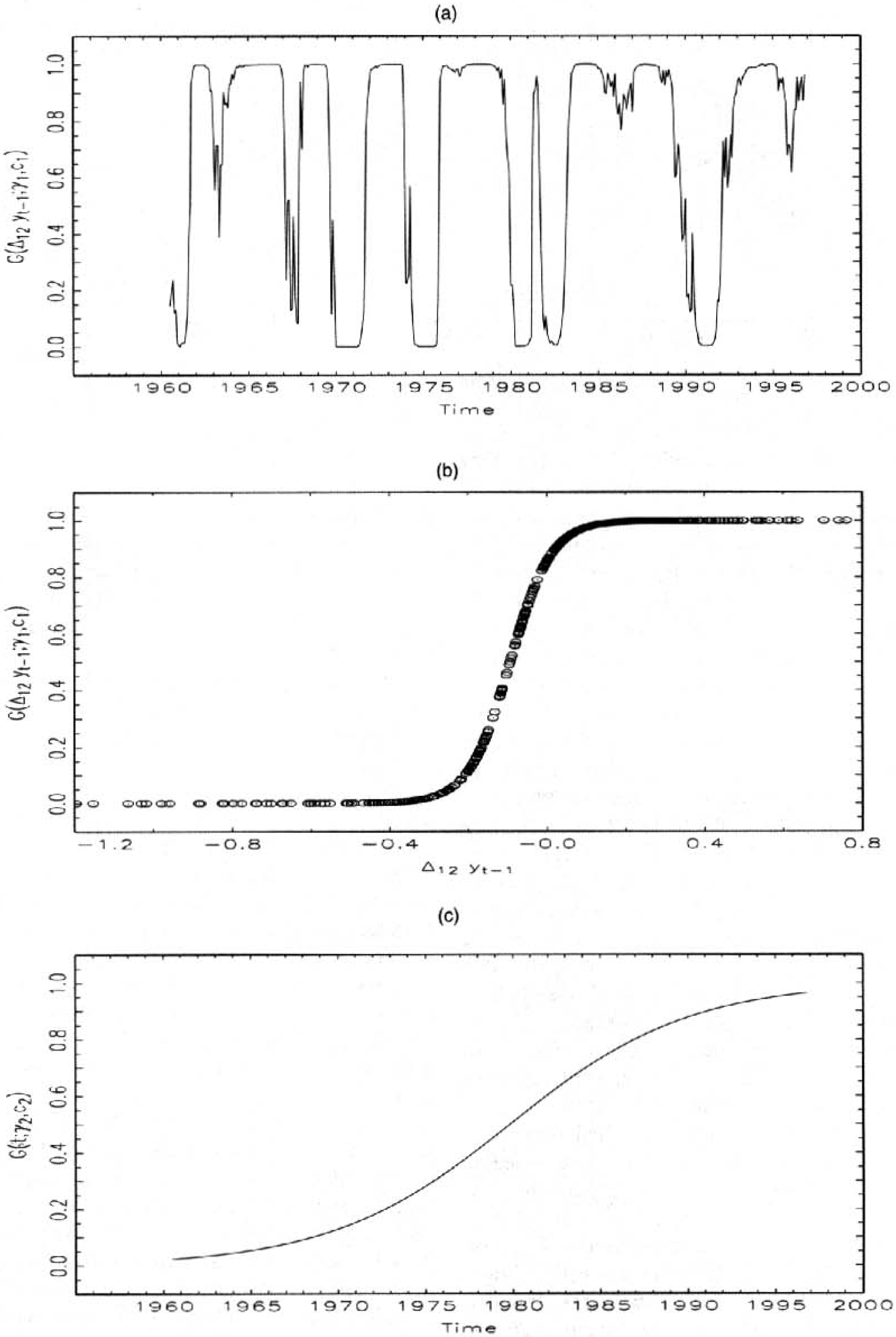

Figure 4. Transition Functions in the TV-STAR Model for the Help-Wanted Index. (a) $G(\Delta_{12}y_{t-1};\gamma_1,c_1) = (1+\exp\{-6.04(\Delta_{12}y_{t-1}+.094)/\sigma_{\Delta_{12}y_{t-1}}\})^{-1}$. *(b)* $G(\Delta_{12}y_{t-1};\gamma_1,c_1) = (1+\exp\{-6.04(\Delta_{12}y_{t-1}+.094)/\sigma_{\Delta_{12}y_{t-1}}\})^{-1}$ *Each circle represents an observation. (c)* $G(t;\gamma_2,c_2) = (1+\exp\{-7.07(t/T-.533)\})^{-1}$.

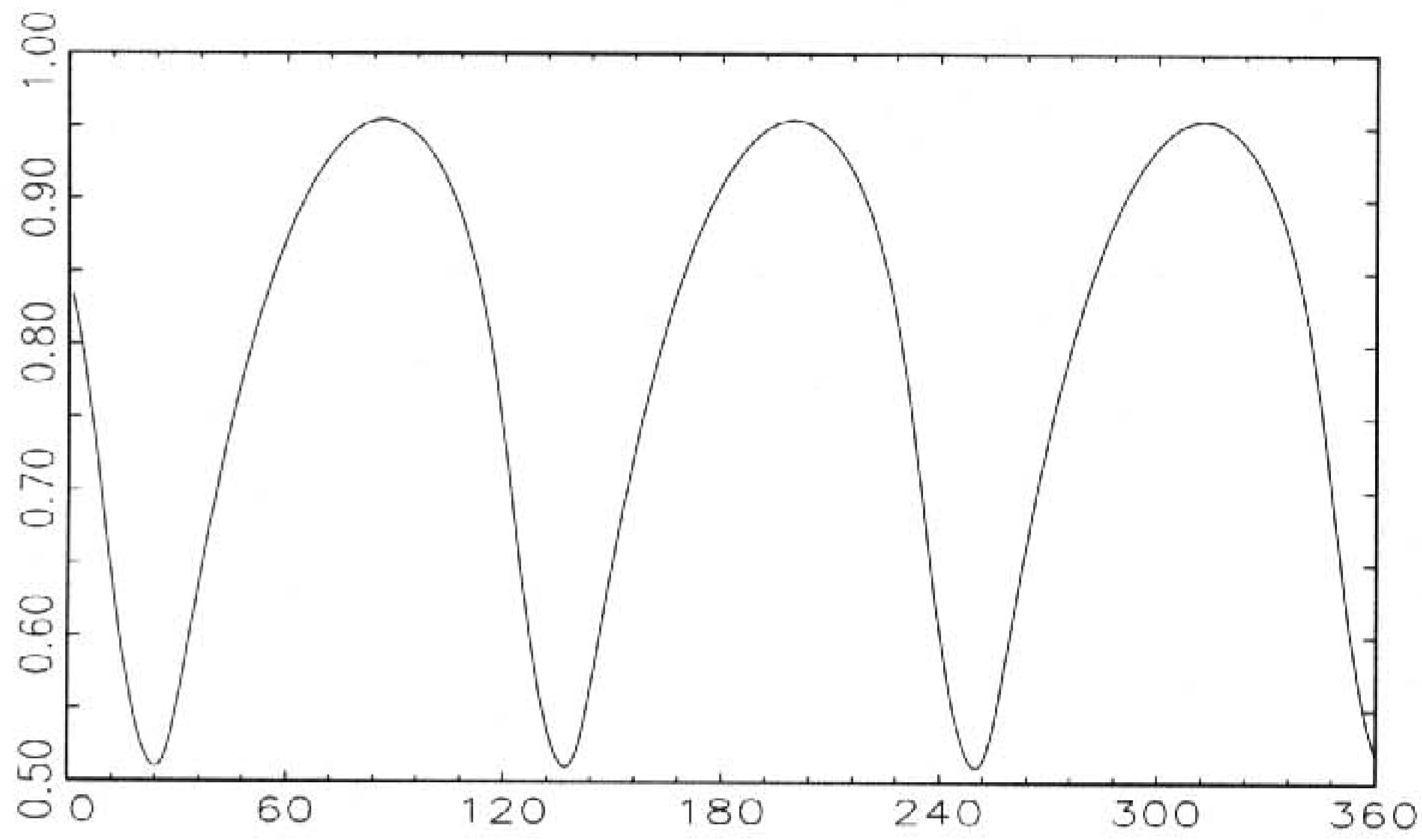

Figure 5. Limit Cycle in the TV-STAR Model for the Monthly Help-Wanted Advertisement Index for $G(t) = 1$.

In applications to physical time series, limit cycles may sometimes have a natural interpretation. In this case it is hardly possible to seriously suggest that the process generating the help-wanted advertising index contains a deterministic cycle. The appearance of the limit cycle may tell us two things, however. First, the persistence of the series has increased over the years. The mean reversion early in the period has given way to a persistent cycle. Second, the shape of the limit cycle indicates that asymmetry is clearly a feature of the series.

To further illustrate the dynamic behavior of our TV-STAR model, we study the persistence of shocks in the model by generalized impulse response functions (GI), defined by Koop, Pesaran, and Potter (1996). To see how the dynamic properties change over time, we do this for the STAR models occurring before the start of the structural change $[G(t) = 0$ in (13)] and after the structural change has been completed $[G(t) = 1]$. To illustrate that the effect of a shock depends on the history of the time series up to the moment that the shock occurs, the GI is computed not only for all histories ("unconditional"), but also separately for those histories for which the value of the transition function $G(\Delta_{12}y_{t-1})$ in (12) is greater ("expansion") and smaller ("recession") than .5. The computational details are given in the Appendix. The GIs are displayed using highest-density regions (see Hyndman 1995, 1996).

Figure 6 shows 50%, 75%, and 90% highest density regions (HDRs) of the different GIs at horizons equal to $0, 3, 6, \ldots, 60$ months. Comparing panels (a) and (b), it is seen that shocks have a stronger effect before the start of the structural change than at the end of the structural change. Before the structural change, the effect of a shock is amplified up to 12–15 months after arrival of the shock, followed by a monotonic decline toward 0. After the structural change, the initial effect is much smaller, whereas the contraction of the density toward 0 is oscillatory. Focusing on panels (c)–(f), it appears that before the structural change, shocks occurring during recessions have larger temporary effects than shocks occurring during expansions. At the horizon of 60 months, the effects are of comparable magnitude once again. After the structural change, the GI for shocks occurring during recessions is larger than the GI for shocks occurring during expansions only for horizons up to 6 months. At longer horizons, the reverse holds. The symmetric shape of the HDRs suggests that there is little asymmetry in the effects of positive and negative shocks, both before and after the structural change.

Finally, we evaluate the gains from allowing for nonlinearity and structural change from a forecasting perspective. We estimate AR, STAR, TV-AR, and TV-STAR models recursively on an expanding window of data, starting with January 1959–December 1984 and extending up until January 1959–November 1996. All models are estimated in unrestricted form, that is, with four lagged first differences and a lagged level term in each of the regimes. For each window, 1-step-ahead to 12-steps-ahead forecasts for the level of the series are formed, rendering a total of 144 1-step-ahead forecasts, 143 2-steps-ahead forecasts, and so on up to 133 12-steps-ahead forecasts. Table 8 contains mean squared prediction errors (MSPEs) for selected horizons. For forecast horizons up to 6 months, the STAR model achieves the smallest MSPE, whereas for longer horizons, the TV-STAR model has the best forecasting performance. For example, compared with the AR model, the MSPE is reduced by 40% for 12-months ahead forecasts. It is noteworthy that at both short and long horizons, allowing only for time-varying parameters leads to more inaccurate forecasts than those obtained from the AR model. Table 9 presents pairwise model comparisons based on Diebold–Mariano statistics for equality of MSPEs and for forecast encompassing, which in general confirm the aforementioned observations. Note that even though the MSPE reduction at the 12-month horizon achieved by the TV-STAR model appears substantial, we cannot formally reject the null hypothesis that the MSPEs of the other models are equal at conventional significance levels.

6. CONCLUDING REMARKS

In this article we have considered a smooth transition model (the TV-STAR model) that allows for regime-switching behavior in conjunction with time-varying parameters. We developed two strategies for model building, specific-to-general

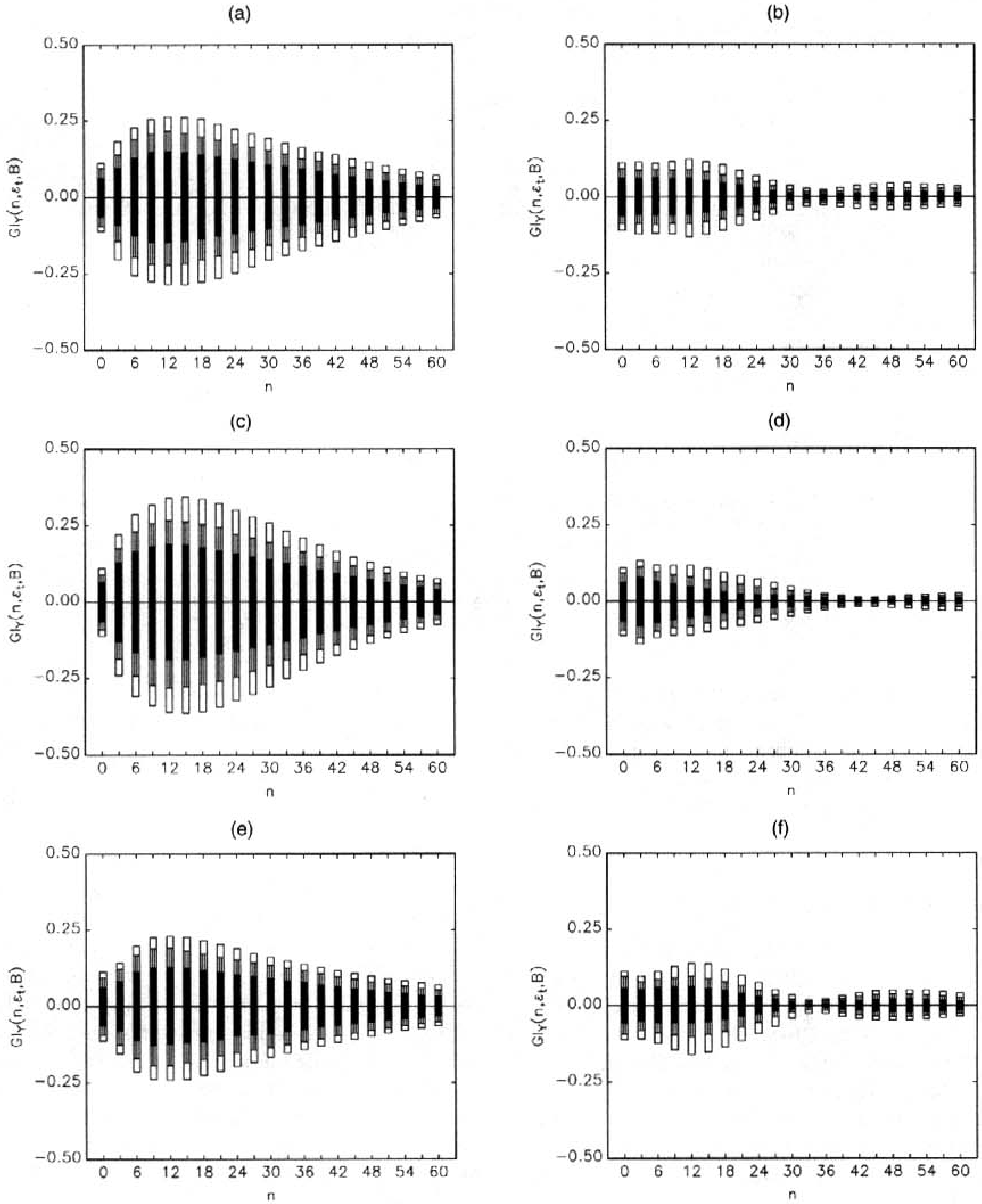

Figure 6. The 50% (Black), 75% (Hatched), and 90% (White) Highest Density Regions for Generalized Impulse Response Functions in the Two-Regime STAR Models for the Help-Wanted Index That Are Effective When $G(t)$ in (13) Equals 0 and 1. Recession and expansion relate to histories for which the value of the transition function $G(\Delta_{12}y_{t-1})$ in (12) is larger and smaller than .5. (a) $G(t) = 0$, unconditional; (b) $G(t) = 1$, unconditional; (c) $G(t) = 0$, recession; (d) $G(t) = 1$, recession; (e) $G(t) = 0$, expansion; (f) $G(t) = 1$, expansion.

Table 8. Mean Squared Prediction Errors of Models for the Help-Wanted Index

Horizon	*AR*	*STAR*	*TV-AR*	*TV-STAR*
1	.93	.88	.91	.90
3	2.15	1.92	2.31	2.23
6	6.42	5.17	7.15	5.48
9	12.01	9.46	13.92	8.50
12	20.12	16.56	22.40	12.11

NOTE: Mean squared prediction errors of levels forecasts from AR, STAR, TV-AR and TV-STAR models for the help-wanted index for the period January 1985–December 1996.

and specific-to-general-to-specific. Monte Carlo simulations showed that the differences in the performance of the two modeling strategies are not great. Because the LM-type tests used in the specific-to-general-to-specific procedure only require estimation of linear models, they can be used to obtain an impression of the importance of nonlinearity and/or structural change in a particular time series. We used this to our advantage when we examined potential nonlinearity and/or structural instability of numerous U.S. macroeconomic time series. The specific-to-general procedure is most useful when it comes to careful specification of a model with nonlinear

Table 9. Forecast Evaluation of Models for the Help-Wanted Index

	MSPE tests				Forecast encompassing tests			
Model	AR	STAR	TV-AR	TV-STAR	AR	STAR	TV-AR	TV-STAR
h = 1								
AR		.078	.29	.32		6.1E−3	.070	9.2E−3
STAR	.92		.73	.60	.60		.14	.064
TV-AR	.71	.27		.41	.34	.011		.014
TV-STAR	.68	.40	.59		.083	.053	.049	
h = 6								
AR		.11	.82	.31		.056	.31	.078
STAR	.89		.95	.61	.79		.37	.076
TV-AR	.18	.047		.21	.032	7.2E−4		8.6E−3
TV-STAR	.69	.39	.79		.21	.12	.19	
h = 12								
AR		.19	.67	.18		.12	.25	.12
STAR	.81		.85	.19	.70		.22	.11
TV-AR	.33	.15		.082	.12	.016		.016
TV-STAR	.82	.81	.92		.48	.28	.40	

NOTE: The table presents pairwise model comparisons of the out-of-sample forecasting results for the AR, STAR, TV-AR, and TV-STAR models for the help-wanted index for the period January 1985–December 1996. The (i, j)th entry in the panels under the heading "MSPE tests" is the p value of the statistic of Diebold and Mariano (1995) for testing the null hypothesis that model i's forecast performance at horizon h as measured by MSPE is equal to that of model j, against the one-sided alternative that the forecast performance of model j is better. The (i, j)th entry in the panels under the heading "Forecast encompassing tests" is the p value of the Diebold–Mariano type forecasting encompassing statistic of Harvey, Leybourne, and Newbold (1998) for testing the null hypothesis that model i's forecast at horizon h encompasses model j's forecast.

and/or time-varying properties. We demonstrated this by the detailed example involving the help-wanted advertising index. In short, the TV-STAR model appears to be a practical tool in investigations that take account of the possibility of simultaneous nonlinearity and parameter instability in economic time series.

ACKNOWLEDGEMENTS

Financial support from the Jan Wallander and Tom Hedelius Foundation for Social Research (Contract No. J99/37) is gratefully acknowledged. The first author acknowledges financial support from the Tore Browaldh Foundation; the second author, from the Swedish Council for Research in the Humanities and Social Sciences. Part of this research was done while the third author was visiting the Graduate School of Business, University of Chicago. The hospitality and stimulating research environment provided are gratefully acknowledged. Material from preliminary versions of this paper was presented at the International Symposium on Forecasting, Edinburgh, June 1998, and the NBER/NSF Time Series Meeting, Chicago, September 1998. Comments from participants at these meetings, as well as from seminar participants at the Cardiff Business School, Loughborough University, University of Nottingham, and University of Warwick, two anonymous referees, and an associate editor, are highly appreciated. Any remaining errors are ours.

APPENDIX: GENERALIZED IMPULSE RESPONSE FUNCTIONS AND HIGHEST DENSITY REGIONS

In this appendix we discuss how we obtain generalized impulse response functions and highest-density regions for the two-regime STAR models that occur when $G(t)$ in (13) is equal to 0 and 1.

The GI for a specific shock $\varepsilon_t = \delta$, and specific information set or "history," $\Omega_{t-1} = \omega_{t-1}$, is defined as

$$GI_{\Delta y}(n, \delta, \omega_{t-1}) = \mathrm{E}[\Delta y_{t+n} | \varepsilon_t = \delta, \Omega_{t-1} = \omega_{t-1}] - \mathrm{E}[\Delta y_{t+n} | \Omega_{t-1} = \omega_{t-1}]. \quad \text{(A.1)}$$

To obtain relevant histories ω_{t-1}, we generate 250 series of 2,500 observations from the two-regime STAR models that are effective when $G(t)$ in (13) is equal to 0 and 1 by sampling with replacement from the residuals of the estimated TV-STAR model. We use the final values of each series as histories in the impulse response analysis. We consider values of the normalized initial shock equal to $\delta/\hat{\sigma}_\varepsilon = \pm 3, \pm 2.8, \ldots, \pm .2, 0$, where $\hat{\sigma}_\varepsilon$ denotes the estimated standard deviation of the residuals from the TV-STAR model. For each combination of history and initial shock, we compute $GI_{\Delta y}(n, \delta, \omega_{t-1})$ for horizons $n = 0, 1, \ldots, N$ with $N = 60$. The conditional expectations in (A.1) are estimated as the means over 1,000 realizations of Δy_{t+n}, obtained by iterating on the STAR model, with and without using the selected initial shock to obtain Δy_t, and using randomly sampled residuals of the estimated TV-STAR model elsewhere. Impulse responses for the level of the help-wanted index are obtained by accumulating the impulse responses for the first differences, that is, $GI_y(n, \delta, \omega_{t-1}) = \sum_{i=0}^{n} GI_{\Delta y}(i, \delta, \omega_{t-1})$.

The GIs for specific histories and shocks are used to estimate the density of $GI_y(n, \varepsilon_t, B)$, where B denotes a set of selected histories. The set B is taken to be either all histories ("unconditional") or those histories for which the transition function $G(\Delta_{12} y_{t-1})$ in (12) is larger ("recession") or smaller ("expansion") than .5. The densities are obtained with a standard Nadaraya–Watson kernel estimator, using $\phi(\delta/\hat{\sigma}_\varepsilon)$ as weight for $GI_y(n, \delta, \omega_{t-1})$, where $\phi(z)$ denotes the standard normal probability distribution. This weighting scheme is used because the standardized shocks $\delta/\hat{\sigma}_\varepsilon$ then are effectively sampled from a discretized normal distribution, and the resulting distribution of $GI_y(n, \varepsilon_t, \Omega_{t-1})$ should resemble a normal distribution if the effect of shocks is symmetric and proportional to their magnitude (as is the case in linear models). Finally, highest-density regions are then estimated using the density quantile method outlined by Hyndman (1996).

[Received April 2000. Revised December 2001.]

REFERENCES

Abraham, K. G. (1987), "Help-Wanted Advertising, Job Vacancies, and Unemployment" (with discussion), *Brookings Papers on Economic Activity*, 18, 207–248.

Andrews, D. W. K. (1993), "Tests for Parameter Instability and Structural Change With Unknown Change Point," *Econometrica*, 61, 821–856.

Andrews, D. W. K., and Ploberger, W. (1994), "Optimal Tests When a Nuisance Parameter Is Present Only Under The Alternative," *Econometrica*, 62, 1383–1414.

Beaudry, P., and Koop, G. (1993), "Do Recessions Permanently Change Output?," *Journal of Monetary Economics*, 31, 149–163.

Birchenhall, C., Jessen, H., Osborn, D. R., and Simpson, P. (1999), "Predicting U.S. Business Cycle Regimes," *Journal of Business & Economic Statistics*, 17, 313–323.

Cooper, S. J. (1998), "Multiple Regimes in U.S. Output Fluctuations," *Journal of Business & Economic Statistics*, 16, 92–100.

Davies, R. B. (1977), "Hypothesis Testing When a Nuisance Parameter Is Present Only Under the Alternative, *Biometrika*, 64, 247–254.

——— (1987), "Hypothesis Testing When a Nuisance Parameter Is Present Only Under the Alternative," *Biometrika*, 74, 33–43.

Diebold, F. X., and Mariano, R. S. (1995), "Comparing Predictive Accuracy," *Journal of Business & Economic Statistics*, 13, 253–263.

Diebold, F. X., and Rudebusch, G. D. (1992), "Have Postwar Economic Fluctuations Been Stabilized?," *American Economic Review*, 82, 993–1004.

Eitrheim, Ø., and Teräsvirta, T. (1996), "Testing the Adequacy of Smooth Transition Autoregressive Models," *Journal of Econometrics*, 74, 59–76.

Granger, C. W. J. (1993), "Strategies for Modelling Nonlinear Time-Series Relationships," *The Economic Record*, 69, 233–238.

Hansen, B. E. (1996), "Inference When a Nuisance Parameter Is not Identified Under the Null Hypothesis," *Econometrica*, 64, 413–430.

Harvey, D., Leybourne, S., and Newbold, P. (1998), "Test for Forecast Encompassing," *Journal of Business & Economic Statistics*, 16, 254–259.

Hyndman, R. J. (1995), "Highest-Density Forecast Regions for Nonlinear and Nonnormal Time Series," *Journal of Forecasting*, 14, 431–441.

——— (1996), "Computing and Graphing Highest-Density Regions," *The American Statistician*, 50, 120–126.

Kim, C.-J., and Nelson, C. R. (1999), "Has the U.S. Economy Become More Stable? A Bayesian Approach Based on a Markov-Switching Model of the Business Cycle," *Review of Economics and Statistics*, 81, 608–616.

Koop, G., Pesaran, M. H., and Potter, S. M. (1996), "Impulse Response Analysis in Nonlinear Multivariate Models," *Journal of Econometrics*, 74, 119–147.

Leybourne, S., Newbold, P., and Vougas, D. (1998), "Unit Roots and Smooth Transitions," *Journal of Time Series Analysis*, 19, 83–97.

Lin, C.-F. J., and Teräsvirta, T. (1994), "Testing the Constancy of Regression Parameters Against Continuous Structural Change," *Journal of Econometrics*, 62, 211–228.

Luginbuhl, R., and De Vos, A. (1999), "Bayesian Analysis of an Unobserved-Component Time Series Model of GDP With Markov-Switching and Time-Varying Growths," *Journal of Business & Economic Statistics*, 17, 456–465.

Lütkepohl, H., Teräsvirta, T., and Wolters, J. (1998), "Investigating the Stability and Linearity of a German M1 Money Demand Function," *Review of Economics and Statistics*, 80, 399–409.

Luukkonen, R., Saikkonen, P., and Teräsvirta, T. (1988), "Testing Linearity Against Smooth Transition Autoregressive Models," *Biometrika*, 75, 491–499.

Parker, R. E., and Rothman, P. (1996), "Further Evidence on the Stabilization of Postwar Economic Fluctucations," *Journal of Macroeconomics*, 18, 289–298.

Pesaran, M. H., and Potter, S. M. (1997), "A Floor and Ceiling Model of U.S. Output," *Journal of Economic Dynamics and Control*, 21, 661–695.

Potter, S. M. (1995), "A Nonlinear Approach to U.S. GNP," *Journal of Applied Econometrics*, 10, 109–125.

Sichel, D. E. (1993), "Business Cycles Asymmetry: A Deeper Look," *Economic Inquiry*, 31, 224–236.

Skalin, J. (1998), "Testing Linearity Against Smooth Transition Autoregression Using a Parametric Bootstrap," SSE/EFI Working Paper Series in Economics and Finance No. 276, Stockholm School of Economics.

Skalin, J., and Teräsvirta, T. (2002), "Modeling Asymmetries and Moving Equilibria in Unemployment Rates," *Macroeconomic Dynamics*, 6, 202–241.

Stock, J. H., and Watson, M. W. (1996), "Evidence on Structural Instability in Macroeconomic Time Series Relations," *Journal of Business & Economic Statistics*, 14, 11–30.

——— (1999), "A Comparison of Linear and Nonlinear Univariate Models for Forecasting Macroeconomic Time Series," in *Cointegration, Causality, and Forecasting: A Festschrift in Honour of Clive W. J. Granger*, eds. R. F. Engle and H. White, Oxford, UK: Oxford University Press, pp. 1–44.

Teräsvirta, T. (1994), "Specification, Estimation, and Evaluation of Smooth Transition Autoregressive Models," *Journal of the American Statistical Association*, 89, 208–218.

——— (1998), "Modeling Economic Relationships with Smooth Transition Regressions," in *Handbook of Applied Economic Statistics*, eds. A. Ullah and D. E. A. Giles, New York: Marcel Dekker, pp. 507–552.

Teräsvirta, T., and Anderson, H. M. (1992), "Characterizing Nonlinearities in Business Cycles Using Smooth Transition Autoregressive Models," *Journal of Applied Econometrics*, 7, S119–S136.

Tiao, G. C., and Tsay, R. S. (1994), "Some Advances in Non-Linear and Adaptive Modelling in Time-Series" (with discussion), *Journal of Forecasting*, 13, 109–140.

Van Dijk, D., and Franses, P. H. (1999), "Modeling Multiple Regimes in the Business Cycle," *Macroeconomic Dynamics*, 3, 311–340.

Watson, M. W. (1994), "Business Cycle Durations and Postwar Stabilization of the U.S. Economy," *American Economic Review*, 84, 24–46.

Wolters, J., Teräsvirta, T., and Lütkepohl, H. (1999), "Modelling the Demand for M3 in the Unified Germany," *Journal of Applied Econometrics*, 14, 511–525.

Wooldridge, J. M. (1991), "On the Application of Robust, Regression-Based Diagnostics to Models of Conditional Means and Conditional Variances," *Journal of Econometrics*, 47, 5–46.

——— (1994), "Estimation and Inference for Dependent Processes," in *Handbook of Econometrics*, vol. IV, eds. R. F. Engle and D. L. McFadden, Amsterdam: Elsevier Science, pp. 2639–2738.

Zagorsky, J. L. (1998), "Job Vacancies in the United States: 1923 to 1994," *Review of Economics and Statistics*, 80, 338–345.